Brown trout and toxic metals: Local adaptation to the legacy of Britain's mining history

Submitted by **Josephine Rosanna Paris** to the University of Exeter
as a thesis for the degree of
Doctor of Philosophy in Biological Sciences
In April 2017

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University

Signature: ……

ABSTRACT

The effect of human activity on the natural world is increasingly shaping the evolution of species. The capacity of evolution to occur in individuals of a species, via natural selection acting on the genotypes of local populations through successive generations, is known as local adaptation. In southwest England, historical mining activity has resulted in a patchwork of highly metal-contaminated rivers across the region. Where the ecological diversity in many of these rivers has been decimated, metal-tolerant brown trout (*Salmo trutta* L.) populations seem to thrive. What are the mechanisms underlying this apparent metal-tolerance? And can it be attributed to processes of local adaptation? This thesis takes a multi-faceted approach in assessing this, by exploring the patterns and processes involved in metal-tolerance in brown trout populations in southwest England. A series of investigations were undertaken, including the use of neutral genetic markers (microsatellites), reduced representation genome sequencing (RAD-seq), common-garden exposure experiments, and genome-wide analysis of hepatic gene expression (RNA-seq). The microsatellite analysis illustrated that metal-tolerant trout have a different genetic architecture compared to fish in clean rivers and, using Bayesian analysis, these demographic differences were correlated with key periods of mining history. We then developed an approach to facilitate robust screening of genome-wide polymorphic loci through a method of parameter optimisation for RAD-seq. This approach formed the basis for identifying loci for investigating the genomic processes of local adaptation in metal-tolerant trout. We present genome-wide (RAD-seq) data highly indicative that neighbouring trout populations, differently impacted by unique 'cocktails' of metal pollutants have evolved both parallel and convergent mechanisms of metal tolerance. Through a common garden experiment, exposing metal-tolerant and metal-naïve fish to a mixture of metals, we were able to hone in on the physiological mechanisms underlying metal-tolerance. Finally, through RNA-seq, we observed that metal-tolerant fish showed little to no changes in hepatic gene expression when exposed to metals, pointing to innate mechanisms of metal handling. Together, the marriage of these various investigations showcases the remarkable ability of local adaptation in conferring metal-tolerance to brown trout populations in southwest England, and, importantly, the resilience of species' in the face of human-altered environments.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I

CHAPTER II

CHAPTER III

CHAPTER IV

CHAPTER V

CHAPTER VI

LIST OF TABLES

CHAPTER V

ACKNOWLEDGEMENTS

There are so many people I would like to thank! This PhD has definitely been a collaborative endeavour, involving umpteen efforts from my supervisors, colleagues, family and friends…

Firstly, I am exceptionally grateful to my supervisor Jamie Stevens, for being a kind friend and an incredibly supportive PhD supervisor. Thank you for your insight, your guidance, and your wise words. You're a true old-school academic, nudging me in the right direction, and providing me with a pick-me-up when you could tell I was crashing. I only hope that one day you'll switch the diet coke for a carbonated water.

Secondly, thanks to my second supervisor Eduarda Santos, for demonstrating dedication for perfection. Thanks for the late-night cups of tea, sweets in your office and suggestions I eat more fruit and get more zzz's. I only wished I'd listened to your words of caution before starting that crazy physiology experiment, but no regrets, two good chapters from that insane summer.

A special thank you to my unofficial supervisor Andy King. You have been an amazing guide, from teaching me how to use a pipette properly, to explaining the numerous idiosyncrasies of population genetics software. Your kindness and mentoring has kept me going and I hugely appreciate everything you've done: answering random phone calls about gene flow, the constant supply of caramel wafers and allowing me to hash out all the details of my project whenever I asked; AND getting excited about it all.

Next, a thank you to all the members of the Molecular Ecology and Evolution Group (MEEG), past and present, for the team support. Especially to Emma Kennedy for too many caramel macchiatos, Charles Ikediashi for countless cups of tea, Andy Bowkett for allowing my naval gazing, Pat Hamilton for the inspiring evolution chats and 'Tjenks' for his persistence with analysis in R, and also to Tracey, Guy and Kirsten.

A big thank you to the funders and stakeholders involved in the project: the Westcountry Rivers Trust, the Environment Agency, and the University of Exeter. A specific acknowledgement to Simon Toms from the Environment

Agency and to Bruce Stockley, Giles Rickard and Matt Healey from the Westcountry Rivers Trust, and again to the WRT guys for organising the permits and the electrofishing for samples, and making collection of Hayle, Red River and Fowey trout possible.

Several people are owed a huge recognition for making many parts of the PhD project feasible: Nic Bury and Rod Wilson for the experimental design and guidance throughout the physiology experiment, as well as the assistance from everyone in the aquarium (especially Darren, Greg and Steve); members of Lab 201 (Anke, John, Lauren, Jenny, Rob and Mauricio); Jan and Phil Shears for the ace electrofishing and also for the cups of tea in their wonderful garden; Konrad and Julian for educating me in bioinformatics, getting me teaching on numerous genomics and RADseq workshops and as a consequence changing the outreach of my research and being a catalyst for many of the close professional relationships I have made.

To my wonderful friends! Special thanks to Emily Cook for ‘all the feels’; Molly Payne for long-term resilience; Krista Sherman for the wine therapy and demonstrating sass; Jane Usher for letting me snot on shoulders and assistance with menial lab-based exercises involving exploding fish tissue; my two close companions Guy Leonard and Sophie Shaw, for all the bioinformatics support and eccentric daily conversations; and Jonathan Green for just being a truly wonderful friend. Also, to the people of The Cave, especially to Shelly, John, Ali, Steph, Victoria, Okhyun and Sara – kitchen parties and kimchi.

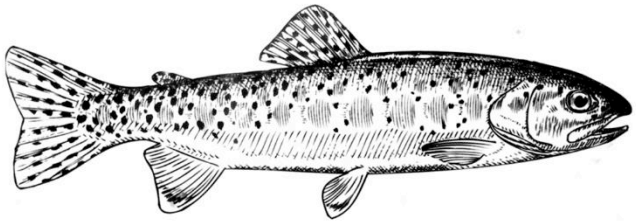
Finally, to my amazing family: my Mum for unfailing love and various Mum activities (sending vitamins in the post, fixing my curtains, and bringing boxes of red wine and purple sprouting broccoli), and the heaps of love from my Dad, my Step-Dad, my Sister and crazy ‘Gma’. Finally, to Martin, for sticking with it and sticking with me. You really kept me going in more ways than I can express gratitude for, and raised a whole lot of laughter in the process.

So long and thanks to all the fish.

LIST OF ABBREVIATIONS

AARC	Atlantic Aquatic Resource Conservation
ABC	Approximate Bayesian Computation
AD	Adriatic mtDNA lineage
AD	anno Domini
AEC	Adaptive evolutionary conservation
AFLPs	Amplified fragment length polymorphisms
Ag	Silver
A_R	Allelic richness
AT	Atlantic mtDNA lineage
BC	Before Christ
BIIS	British-Irish Ice Sheet
BGS	British Geographical Survey
cDNA	Complementary DNA
CI	Confidence interval
cm	Centimetre
CNVs	Copy number variants
Cu	Copper
CWDML	Cornwall and West Devon Mining Landscape World Heritage Site
DA	Danubian mtDNA lineage
DFTD	Devil facial tumour disease
DNA	Deoxyribonucleic acid
EA	Environment Agency
ESUs	Evolutionary Significant Units
FIGCS	Fish Gill Cell System
F_{IS}	Inbreeding coefficient
F_{ST}	Fixation Index
FTE	Full Time Employment
Gly	Glycerate dehydrogenase
GO	Gene Ontology
GWAS	Genome wide association studies
HCT	Hematocrit
H_E	Expected heterozygosity
H_O	Observed heterozygosity
HSI	Hepatosomatic Index
HWE	Hardy-Weinberg Equilibrium
IBA	Isolation-by-adaptation
IBD	Isolation-by-distance
IUCN	World Conservation Union

K	Thousand
kg	Kilogram
km	Kilometres
LD	Linkage disequilibrium
LFMM	Latent Fixed Effect Models
LGM	Last Glacial Maximum
m	Metres
Ma	Million years ago
MA	Marmoratus mtDNA lineage
MAS	Marker-assisted selection
MHC	Major histocompatibility complex
mRNA	Messenger RNA
MT	Metallothioneins
mtDNA	Mitochondrial DNA
MUs	Management Units
MY	Million years
N_E	Effective population size
NERC	National Environment Research Council
PAHs	Polyaromatic hydrocarbons
PGI	Phosphoglucose Isomerase
PCBs	Polychlorinated biphenyls
QTLs	Quantitative Trait Loci
RAD-Seq	Restriction site Associated DNA Sequencing
RAPDs	Random amplified polymorphic DNA
RFLPs	Restriction fragment length polymorphisms
RIGS	Regional Important Geological Sites
RNA	Ribonucleic acid
RNA-Seq	RNA Sequencing
RT-qPCR	Reverse-transcriptase quantitative-Polymerase Chain Reaction
SAGE	Serial Analysis of Gene Expression
SNP	Single nucleotide polymorphism
SSSI	Sites of Special Scientific Interest
SW	Southwest
TEER	Transepithelial electrical resistance
UK	United Kingdom
UNESCO	United Nations Educational, Scientific and Cultural Organisation
WGD	Whole genome duplication



Chapter I

GENERAL INTRODUCTION

To yank me clear takes the sudden, strong spine

Of one of the river's real members –

Thoroughly made of dew, lightning and granite

Very slowly over four years. A trout, a foot long,

Lifting its head in a shawl of water,

Fins banked stuff like a trireme

It forces the final curve wide, getting

A long look at me. So much for the horror

It has changed places.

Ted Hughes

'Growth of the human population and depletion of the earth's natural resources continues unabated.'

E.O. Wilson

Humankind has exploited the world's resources for thousands of years, its appetites increasing in line with a burgeoning and increasingly technologically focused population. The processes involved continue to drive dramatic changes in, *inter alia*, species distributions, niche occupation and local and global patterns of extinction and (re)colonisation. These changes must be studied and documented in a scientific framework if potential harm is to be diminished.

The Convention on Biological Diversity was inaugurated at the Earth Summit in Rio de Janeiro in 1992. Its three main goals were to: (i) conserve biological diversity; (ii) to use the components of biological diversity in a sustainable way; and (iii) promote fair and equitable division of the benefits derived from genetic resources. According to the Convention, if these aims are to be achieved, it is essential to conserve biodiversity at all biological levels. This includes populations harbouring unique genetic mechanisms conferring genetic adaptation and phenotypic plasticity to specific environments (Evolutionary Significant Units, or ESUs).

This thesis investigates a small component of this, by exploring the patterns and processes of metal tolerance observed in local populations of brown trout (*Salmo trutta* L.) in southwest England. These populations are able to withstand extremely high concentrations of toxic metals present in water bodies affected by historic mining activity. Exploring these patterns on small spatial scales offers an applicable model showing how global biodiversity patterns have been affected by human-driven environmental change, and how some species demonstrate a remarkable capacity to adapt to these changes.

1.1 BROWN TROUT BIOLOGY AND ECOLOGY

Brown trout (*Salmo trutta* L.) is a ubiquitous, generalist species with incredibly diverse ecological and life history characteristics. The species is primarily European; its distribution ranges from Ireland in the west to the Ural mountains of Russia in the east, occurring as far north as Iceland and extending south to the Atlas mountains of North Africa (Figure 1), making it the most widely distributed freshwater fish native to the Palearctic region (Bernatchez 2001). Brown trout is a teleost fish of the Salmonidae family, and is more closely related to the Atlantic salmon (*Salmo salar*) than other species bearing the common name 'trout'. These other species belong to the northern Pacific *Oncorhynchus* genus, which includes the rainbow or steelhead trout (*Oncorhynchus mykiss*) and the cutthroat trout (*Oncorhynchus clarkii*).

Figure 1. Distribution of brown trout (*Salmo trutta* L.) across its native range. Adapted from (Jonsson & Jonsson 2011).

In the British Isles, brown trout is one of the most widespread fish species. It occupies approximately 70% of a given river-length, increasing to 80% in some regions such as the southwest of England (Environment Agency 2008a). As a result of human introductions, the species is also present elsewhere in the world. It has been introduced in at least 24 countries (Elliot 1994), where in some cases it is known to adversely affect native salmonid populations (Fausch & White 1981; Verspoor 1988) and other fish species (Townsend & Crowl 1991; Crowl *et al.* 1992).

1.1.1 HABITAT

Brown trout occupy various water bodies including lakes, streams and rivers. Primary habitat requirements are determined by the quantity and quality of flowing water, as well as composition of the riverbed material (Crisp, 2000). Generally, brown trout prefer fast-flowing, well-oxygenated waters, although habitat preference within a water body is determined by life stage, with larger brown trout preferring the deepest areas and smaller individuals occupying shallow riffles at depths between 5 and 30cm (Jonsson & Jonsson 2011). Trout favour cold waters between 13°C and 16°C (Elliot & Elliot 2010), with temperature being the primary factor determining the rate of growth and development (Brown 1946; Elliot 1976; Edwards *et al.* 1979). Bed sediment characteristics are also influential: well-oxygenated gravel beds are essential for spawning (Hogan & Church 1989; Hogan *et al.* 1998), as is a relatively high velocity of near-bed water due to high oxygen consumption by eggs (Wickett 1954; Greig *et al.* 2007).

Brown trout are opportunistic feeders, their diet varying depending on life stage, local availability of resources and season (Bridcut & Giller 1995; Hyvarinen & Huusko 2006). However, they primarily feed on aquatic and terrestrial invertebrates but may also adopt a piscivorous habit (L'Abee-Lund *et al.* 1992; Jonsson *et al.* 1999, Jensen *et al.* 2012). They play an important role in riverine ecosystems due to their high trophic position (McHugh *et al.* 2008), and are known to alter food-web composition; for example, the absence of trout is associated with significantly higher densities of invertebrate predators (Meissner & Muotka 2006).

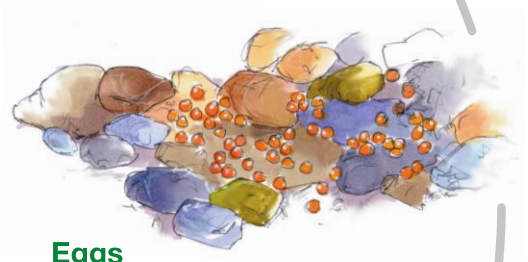
The Brown Trout Life Cycle

Spawning

Between November and February a female brown trout digs a nest or redd in gravel shallows. As she releases her eggs, they are fertilised by the male and then covered with gravel. The gravels must be 10-40mm in size, loose and free of silt with plenty of oxygen rich water flowing through them.

Adults

Adult trout have a territory that gives them a good supply of food and a place to hide from predators, preferring deeper pools. In winter, they migrate, perhaps miles upriver, to spawn. Brown trout live up to 5-20 years.



Eggs

Eggs, 2-5mm in diameter, hatch into alevins in a few months, depending on temperature.

Parr

Fry and parr are territorial and solitary. They need plenty of cover in the river from stones, weed and trailing bankside plants, and shallow water that is not too fast flowing. Only around 5% of young trout survive their first year of life.

Alevins

Alevins stay in the gravel, living off the yolk sac. They then emerge as fry, set up territories and grow into parr.

Trout illustrations by DAB graphics.
www.wildtrout.org

Figure 2. Basic lifecycle of a brown trout. Kindly provided by The Wild Trout Trust (http://www.wildtrout.org/)

1.1.2 LIFE CYCLE

Fork length of a resident brown trout at maturity ranges between 20 to 40cm (King *pers comm.*). However, growth and body mass vary hugely with shifts in biotic and abiotic factors, both within and between rivers (e.g. Vøllestad *et al.* 2002; Parra *et al.* 2011). The average generation time is between 3.5 to 4 years (Hansen *et al.* 2002; Jensen *et al.* 2008).

A typical brown trout lifecycle begins between November and February, when the water temperature is low (0-13°C) and well oxygenated (Environment Agency, 2008). Female trout excavate a shallow hollow in a gravel bed (a redd) into which they lay pockets of eggs; the eggs are then fertilised by one or more males. During the intragravel stage, lasting a few months, embryos develop into alevin and are reliant on the resources of the yolk sac until the young fish, known as fry, emerge in the spring (March-May, highly dependent on temperature; Elliot & Elliot, 2010). Fry disperse from the redd site, establish territories and feed until they become parr. Parr remain in the river for 1 to 3 years (designated 0+ parr, 1+ parr and 2+ parr) until becoming sexually mature adults.

1.1.3 LIFE HISTORY STRATEGIES

Variation on the basic lifecycle of brown trout is incredibly wide. After maturing into parr, a trout may smolt and migrate to sea, spending between a few months to two years feeding, before returning to its natal river to spawn. These anadromous fish are known as sea trout. This anadromous lifestyle is itself also very plastic, with variations in migration patterns (e.g. Lacustrine migration, Denic & Geist 2009), the number of years spent at sea (L'Abee-Lund *et al.* 1989) and in the proportion of trout that migrate within a river system (Jonsson 1985; Aarestrup & Jepsen, 1998; Goodwin *et al.* 2016). Both genetic mechanisms (Gyllensten, 1985; Nichols *et al.* 2008) and environmental cues (Byrne *et al.* 2004; Olsson *et al.* 2006) contribute to variation in migration strategy, and more recent analyses have identified regions of the genome that affect the propensity to migrate in other salmonid species (Hecht *et al.* 2013; Johnston *et al.* 2014; Barson *et al.* 2015). It should also be noted that many freshwater or resident brown trout also undergo some form of within-catchment migration during their lifetime (Ferguson 2006).

The brown trout species therefore occurs as two morphotypes: the resident brown trout, which remains within the river for the duration of its life, and the anadromous sea trout, which migrates to sea for a period of time. Like Atlantic salmon, a key behavioural feature of anadromous trout is the ability to home to their natal spawning site (Stuart 1957; Stabell 1984), a process that is influenced, and can therefore be disrupted, by behavioural, chemical and physical cues (Tierny *et al.* 2010; Maryoung *et al.* 2015). The accuracy of natal homing can be very high (94-98%; Garcia de Leaniz *et al.* 2007), but other data suggest that high rates of straying in returning adults in some populations may be common (15.5%: Berg & Berg 1987; >10% King *et al.* 2016).

The biology of the sea trout lifecycle strategy will not be examined in great detail here, but it is important to note that the straying of anadromous individuals into neighbouring rivers is the main contributor to gene flow and recolonisation in salmonid species (see section 1.1.4) and significantly contributes to migration between rivers and catchments (McKewon *et al.* 2010; Griffiths *et al.* 2011; Ikediashi *et al.* 2012).

1.1.4 PHYLOGENETICS AND PHYLOGEOGRAPHY

The Salmonidae family belong to the large and diverse clade of ray-finned or teleost fish (Teleostei), which is by far the most species-rich vertebrate clade, likely due the 'Teleost-Specific Whole Genome Duplication' (Figure 3; Glasauer & Neuhauss 2014). The ancestors of the Salmonidae family likely appeared between 63 million and 135 million years ago (Lengendre, 1980), with recent estimates placing the origin of the family at 59.1 million years (CI: 63.2-58.1 MY; Crête-Lafrenière *et al.* 2012). The family is defined by an ancient whole genome duplication (WGD) event that occurred between 20 to 120 million years ago (Figure 3; Ohno 1970; Allendorf & Thorgaard 1984; Johnson *et al.* 1987). Extant salmonid species are therefore considered pseudo-tetraploids whose genomes are in the process of reverting to a stable diploid state (Quinn *et al.* 2010 Lien *et al.* 2011).

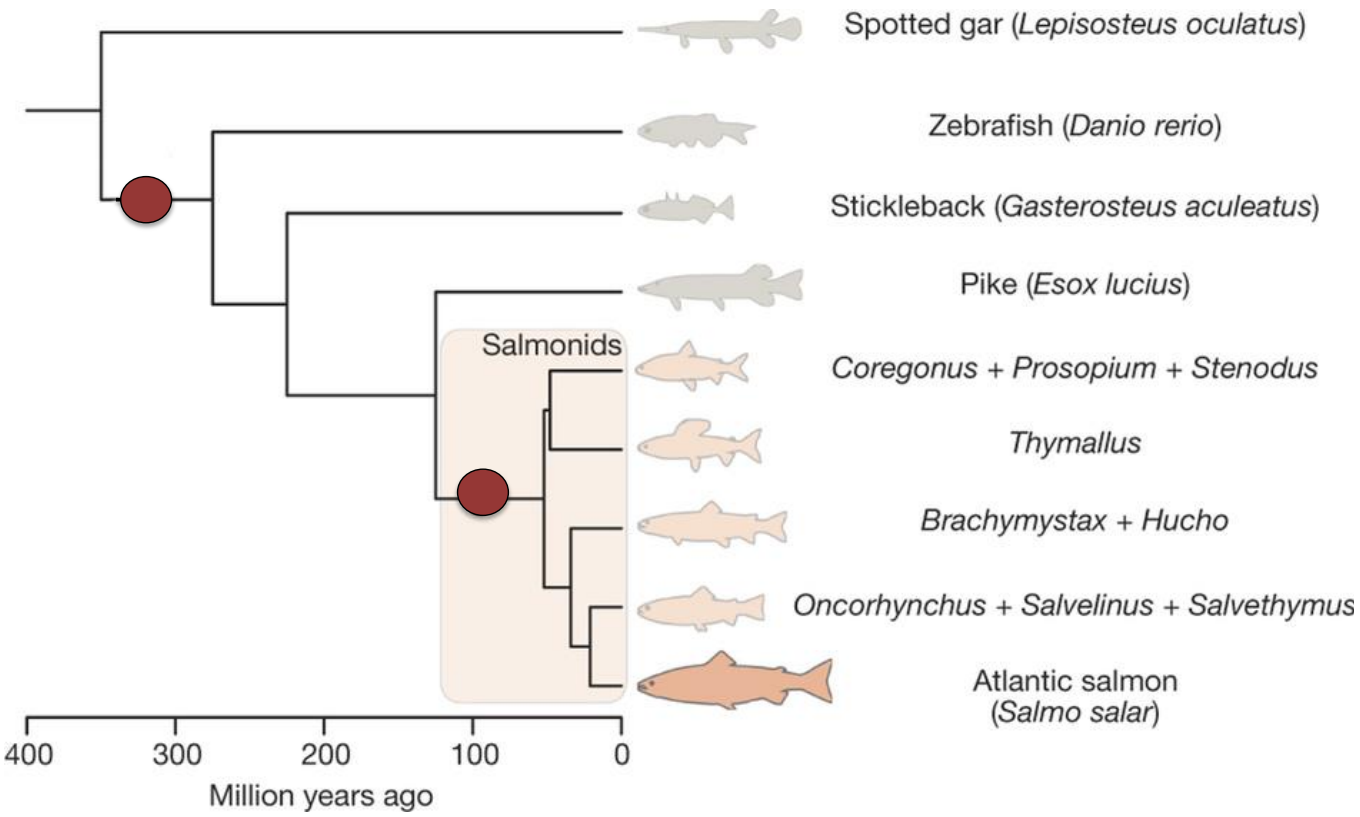


Figure 3. Phylogenetic tree of the salmonid family. Adapted from Lien *et al.* (2016). Dark red circles represent whole-genome duplication (WGD) events.

The species-rich Salmonidae family includes: the whitefishes (*Coregonus*, *Prosopium* and *Stenodus* genera); graylings (*Thymallus*); the lenoks or Asiatic trout (*Brachymystax*) and the *Hucho* genus; the sub-family Salmoninae, which includes the charr (*Salvelinus* and *Salvethymus* genera); the Pacific salmon and trout (*Oncorhynchus* genera); and the North Atlantic salmon and trout – *Salmo* – the Atlantic salmon *Salmo salar* and the brown trout *Salmo trutta.* Brown trout and Atlantic salmon appear to be closely related according to both morphological and molecular data (Elo *et al.* 1997; Crespi & Fulton 2004); the two species are known to hybridise (Vuorinen & Piironen 1984; Verspoor 1988), and gene synteny between the two species is conserved (Moore *et al.* 2005; Gharbi *et al.* 2006).

Across its native range, five major mtDNA brown trout lineages have been identified (Bernatchez 2001). These diverged between 450,000 to 700,000 years ago and consist of: (i) the Atlantic basin & Morocco (AT); (ii) Danubian (DA); (iii) Mediterranean (ME); (iv) marmoratus (MA), associated with the marbled phenotype of the Adriatic Sea and potentially representing a different species (*Salmo marmoratus*); and (v) Adriatic (AD). Admixture is known to occur between these five main evolutionary lineages (Sanz Ball-Llosera *et al.* 2002; Meraner *et al.* 2010; Berrebi 2015). This picture has been complicated by the identification of a possible sixth lineage from the Iberian peninsula (Suárez *et al.* 2001; Martinez *et al.* 2007; Bouza *et al.* 2008) and, more recently, evidence of significant admixture and conflated distribution patterns of the AT and DA lineages (Lerceteau-Köhler *et al.* 2013; Schenekar *et al.* 2014). Nonetheless, British brown trout sit within the AT lineage, which split from the eastern *Salmo trutta* groups approximately 700,000 years ago (Bernatchez 2001). Across Britain, the AT lineage is represented by a finer phylogeographic structure consisting of 25 brown trout mtDNA RFLP (restriction fragment length polymorphism) haplotypes (McKeown *et al.* 2010).

One of the primary forces driving contemporary distributions and spatial structuring of species are glacial ice ages. The Pleistocene glaciation (~2.58Ma – present) has been marked by several glacial and interglacial periods (Lee *et al.* 2011). The most recent of these glaciation events is referred to as the Last Glacial Maximum (LGM), in which glacial retreat began around 19-20K years ago (Clark *et al.*, 2009). At the height of the LGM, species present in glaciated areas either became extinct (Shapiro

et al. 2004; Jacobi *et al.* 2009; Lister & Stuart 2008), retreated away from the ice, or survived in refugia from which they subsequently dispersed.

The majority of Britain was covered with ice during the LGM, rendering the survival of brown trout populations unlikely. Evidence suggests that postglacial colonisation occurred via genetically distinct lineages of anadromous sea trout around 14,000 years ago (Ferguson 1989). Several refugia for these lineages have been posited for brown trout (García-Marín *et al.* 1999; Sanz *et al.* 2000; Weiss *et al.* 2000). British trout most likely radiated from a southern refugium located south of Britain and western France; such refugia have previously been identified for Atlantic salmon in the Iberian Peninsula (Consuegra *et al.* 2002) and northwest France (Finnegan *et al.* 2013), and it would seem probable that these were also utilised by brown trout. These refugia were themselves likely populated by a mixture of extant genetic clades, but ancestral mtDNA haplotypes have been found across Britain and Ireland, showing a high frequency of >75% in the southwest of England (McKewon *et al.* 2010). This area, as well as southwest Ireland, would have been the first region to be colonised, assuming northerly movement from a southern refuge is supported.

Interestingly, the British-Irish Ice Sheet (BIIS) during the LGM did not extend as far south as southwest England (Chiverrell & Thomas 2010), suggesting that relic southwest populations could have survived the LGM in a periglacial refugium (Coard & Chamberlain 1999; Kelly *et al.* 2010). To date, there has been no conclusive evidence to support the theory of a southwest refugium for brown trout (Griffiths, 2005) and further research would be required to confirm this.

The contribution of stochastic phylogeography should be considered in the context of contemporary population genetic structure. Today, rivers in the south of England primarily comprise small catchments: those in the southeast discharging into the English Channel and Straits of Dover and southwest catchments (forming part of the 'Southwest Approaches') draining into the Celtic Sea, the Bristol Channel and the English Channel. Populations of brown trout occupying these southwest rivers are primarily characterised by patterns of strong genetic structuring, which vary between river-basins, and accompanying patterns of isolation-by-distance (Griffiths *et al.* 2009; King *et al*. 2016; King & Stevens, *unpublished data*). This contemporary population structure is likely more influenced by recent demographic processes such

as genetic drift, effective population size reductions and local adaptation of each population to its river environment (see section 1.3.3.3).

1.1.5 SOCIO-ECONOMIC VALUE

Brown trout are of considerable interest to both commercial fisheries and recreational sports fishing. In England and Wales alone, trout angling supports 5,600 jobs and generates approximately £150 million in household incomes annually, mostly in relation to local activity (Environment Agency 2009). It has been estimated that if trout angling were to discontinue, there would be losses of approximately 1,600 jobs and £40 million to the economy (Environment Agency 2009). In the southwest of England, fishing inland waters totalled about £100 million, supporting approximately 2,300 jobs and £50 million of household income in the region (Environment Agency 2009). In northeast Scotland, it was estimated that the combined value of salmon and sea trout fisheries contributed £11.6 million per annum (Butler *et al.* 2009); anglers spent £14.7 million across Scotland as a whole, £4.2 million of which was spent by visiting anglers (Radford & Gibson, 2004). Brown trout are of lesser value in aquaculture, with an estimated global production of just under 5,000 tonnes in 2010 (Frank-Gopolos *et al.* 2015).

1.1.6 CONSERVATION STATUS AND THREATS

Brown trout are designated 'Least Concern' by the World Conservation Union (IUCN; Freyhof 2011) but, nonetheless, the species faces numerous conservation and population viability threats, especially with regard to local populations (Fergurson 1989, Almodovar *et al.* 2006). Trout are only lightly exploited by direct fishing (Braña *et al.* 1990; Almodovar & Nicola 2004) and the main contribution to changes in local trout abundance and diversity are caused by damage, disruption and pollution of the environments and habitats on which the species depends (Crisp 2000).

Modification of river habitats, often in the form of changes in sediment regimes, commercial forestry and flow regulation (Scottish Environment Protection Agency 2002) have altered the availability of suitable habitat (Cattanéo *et al.* 2002; Scheurer *et al.* 2009). Pollution has had a huge effect on fish populations (Hamilton *et al.* 2016), including metal pollution (Larsson *et al.* 1985; Norris *et al.* 1999), agricultural run-off (Uren Webster & Santos 2015; Uren Webster *et al.* 2015), eutrophication

causing hypoxia (Pollock *et al.* 2007), polychlorinated biphenyls (PCBs) (Johansson *et al.* 1972; Belpaeme *et al.* 1996; Ondarza *et al.* 2011) and polyaromatic hydrocarbons (PAHs) (Luckenbach *et al.* 2003; Meland *et al.* 2010). There is also concern that various chemicals such as pharmaceuticals (Brown *et al.* 2014) and nanoparticles (Scown *et al.* 2010; van Aerle *et al.* 2013) may negatively affect natural fish populations**.** Significant disruption to gene flow has been caused by habitat fragmentation (Gosset *et al.* 2006) and artificial barriers to migration, such as dams and weirs (Thorstad *et al.* 2008). Stocking of rivers with hatchery-reared populations has been reported to cause loss of local structure and population viability (Hansen *et al.* 2001; Sanz *et al.* 2006).

In the global context, salmonids, being so heavily reliant on cold waters, are particularly susceptible to climate change (Jonsson & Jonsson 2009; Almodovar 2012). Although brown trout are more resilient to low pH compared to Atlantic salmon (Alabaster & Lloyd 1982), extreme water acidification is also a threat. Research in Norway indicated that by 1983, the effects of acidification had extirpated brown trout from 1,631 lakes, with some regions losing close to 70% of their local populations (Owen *et al.* 2012). For sea trout, stocks are also faced with threats at sea, the primary factor being parasitism by the sea louse *Lepeophtheirus salmonis*, which causes significant physiological disruption (Wells *et al.* 2006), and death (Birkeland 1996). Infestation is likely influenced by proximity to salmon farms (MacKenzie *et al.* 1998; Costello 2009; Butler & Watt 2003). Significant threats to sea trout also include local and commercial net fisheries, such as the Northeast Drift Net Fishery, which was estimated to have caught over 40,000 sea trout in 2013 (Salmon & Trout Conservation UK).

In summary, the biology of brown trout is incredibly dynamic and is heavily influenced by biotic and abiotic factors. While brown trout are vulnerable to the effects of ecosystem change in the natural habitats on which they depend, the plasticity of the species and its resilience to changing conditions in the environment contribute to its success as a global species.

1.2 GEOLOGY AND METAL MINING IN SOUTHWEST ENGLAND

In the context of this thesis, southwest England refers to the counties of Cornwall and Devon. For its size, southwest England has one of the most diverse ranges of minerals in the world (Scrivener *et al.* 1997; Camm 2010). As a result, the area has experienced a long history of mining and smelting activity, primarily focused upon the exploitation of tin (Sn), copper (Cu) and arsenic (As) (West *et al* 1997). During the height of the mining industry in the 18th and 19th Centuries, southwest England was the 'Silicon Valley of its day' (BBC). In 2006, the area was designated as the UNESCO 'Cornwall and West Devon Mining Landscape World Heritage Site' (CWDML) and many other parts of the region have Sites of Special Scientific Interest (SSSI) and Regional Important Geological Sites (RIGS) status.

Across the UK, the largest discharges of metals into rivers and the sea come from abandoned mines (Environment Agency 2013), with an estimated 400 km of the rivers in England and Wales being affected by minewater (Pentreath 1994). The problems of metal-contamination are particularly pronounced in the CWDML region; although the last active mine (South Crofty) was closed in 1998, metal contamination caused by the mining activity persists and remains a major problem. Drainage from these mines over such long periods of time has led to exceedingly high levels of metals in several rivers across the region (Bryan & Hummerstone 1977; Pirrie *et al.* 1997; Pirrie *et al.* 1999; Pirrie *et al.* 2000). Evidence suggests that >70% of deposited contaminants can remain within a river system for >200 years after mine closure (Coulthard & Macklin 2003).

1.2.1 GEOLOGY

The history of mining for metals in the region begins with its diverse geology – an underlying bedrock rich with mineral resources. The area's primary rock composition is sedimentary, formed of Devonian rocks and Carboniferous-age Culm Measures (Stanier 2011). Devonian rocks were formed approximately 400 million years ago, with Carboniferous rocks forming later, around 350 million years ago. This geology is associated with low-alkalinity and acidic waters (Pickering & Owen 1997).

Against the backdrop of Devonian-Carboniferous sedimentary geology are the igneous granite intrusions of the Isles of Scilly, the Land's End Peninsula/Penwith, the Carnmenellis granite, St. Austell Downs, Bodmin Moor, and Dartmoor (Dines, 1956), known collectively as the 'Cornubian batholith' (Figure 4). These granite intrusions are particularly rich in metal resources. During the intrusion of the granite magma, hydrothermal activity leached metallic elements from the surrounding rock, creating the wealth of metals found in ore deposits, or 'lodes' (Camm 2010), and weathering of these lodes over time exposed the metalliferous material. Natural geological processes define the deposition of metals in these lodes: tin is deposited at the bottom, followed by copper, zinc, lead and iron (Dines 1956; Stanier 2011, Camm 2010). Associated with these main layers of metal deposition are other metals such as tungsten and arsenic; mining for a metal of interest typically releases many other metals into the environment.

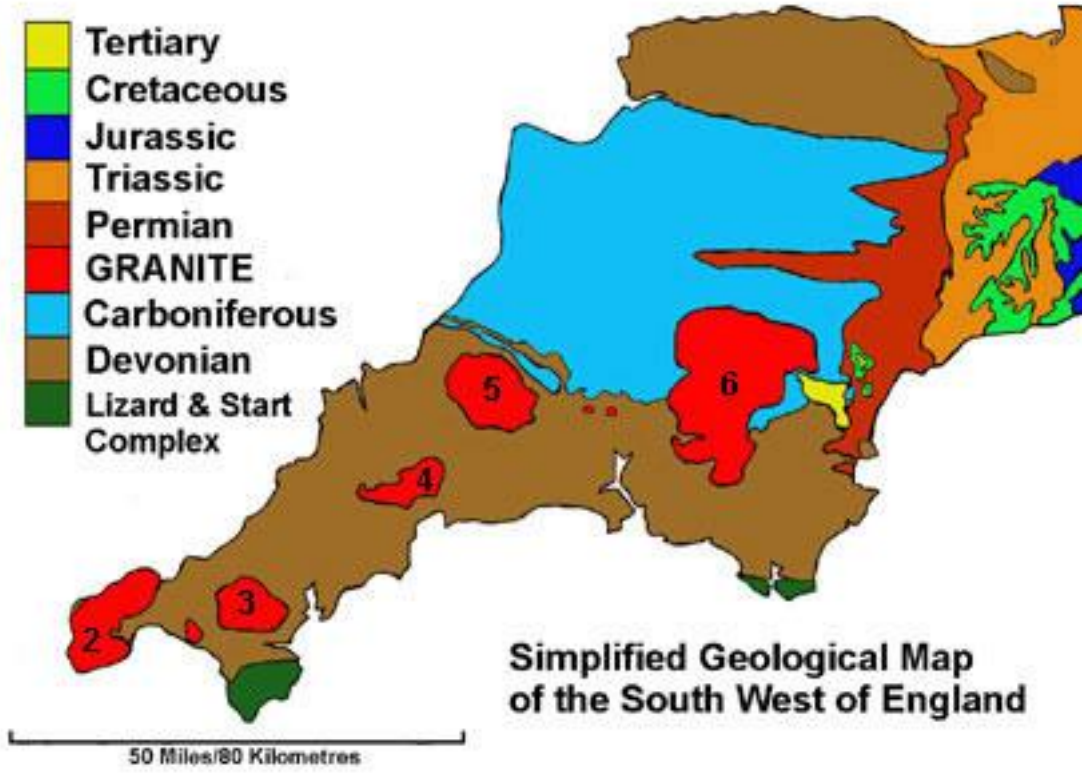


Figure 4. Basic map of the geology of Cornwall and Devon (from summitpost.org). Granite intrusions of the 'Cornubian batholith' are numbered: 1. Isles of Scilly; 2. Land's End; 3. Carnmenellis granite; 4. St. Austell Downs; 5. Bodmin Moor; 6. Dartmoor. These areas are associated with a more metal-rich geology.

It is important to note that several rivers across southwest England drain from these granite intrusions and therefore bear naturally mineralised waters, boasting a metal-rich chemistry (Stanier 2011). The pH of groundwater surrounding granite is generally more acidic than that of the surrounding Devonian-Carboniferous sediment (as low as pH 5.7; Smedley 1991). Although local variation exists in the geochemistry of the water, transition metals, such as copper and zinc, are soluble in acidic mine waters (Bowell & Bruce 1995). Rivers that do not experience contemporary metal pollution from mine wastewater may, therefore, contain ambient background levels of metals and may also have been negatively affected by various mining practices. Throughout this thesis, rivers that experience lower natural concentrations of metals are referred to as 'clean'; these may, however, still harbour higher concentrations compared to rivers not lying on metalliferous bedrock. Therefore, these rivers are designated 'clean' only in comparison to 'metal-impacted' rivers that experience considerably higher contemporary metal pollution.

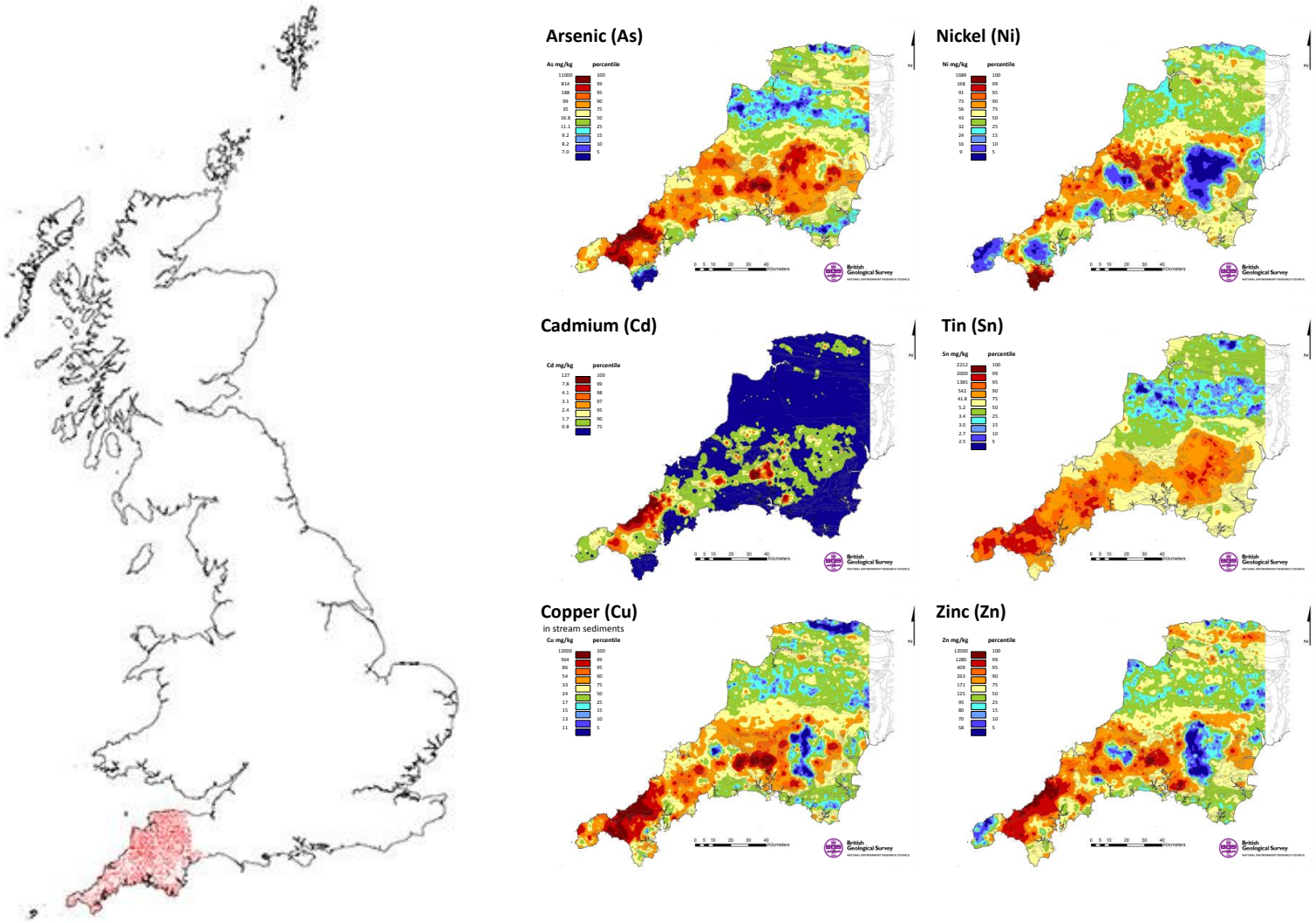


Figure 5. Geochemical baseline maps of Southwest England based on stream sediment sampling provided and adapted from the G-BASE Project. Contains British Geological Survey materials © NERC [2016]. Intense metal contamination is associated with granite intrusions.

1.2.2 PRE-INDUSTRIAL MINING

In order to fully appreciate the selection pressures exerted upon brown trout by mining activity, it is necessary to recount the main periods in history during which these pressures would have been at their greatest.

The history of mining extends back at least 4,000 years to the Bronze Age (c.2500-c. 800 BC) (Hatcher 1973; Timberlake 2001). The tin industry was established in the region by the Early Bronze Age (c. 2100-1500 AD) and was widespread by the Middle Bronze Age (1500-800 AD) (Buckley 2012). Tin is one of the oldest English exports, with Cornwall and Devon possessing a virtual monopoly over European tin production for almost a millennium (Hatcher 1973), although copper, lead, arsenic and zinc have also been extracted on a large scale, with smaller-scale operations extracting a variety of other metals including silver and uranium (Rollinson *et al.* 2008). The copper, lead and zinc industries developed later than the tin industry; systematic mining for copper in the region does not appear to have commenced before the 16th century (Dines 1956).

It has been suggested that the metal resources of England were a primary reason for the Roman invasion (Goody 2012). The Romans were likely the first civilization to mine copper ores in Britain (Collingwood & Myres 1937); indeed, there is substantial documentary evidence of Roman mining, particularly for tin (Manning 1979) but also for copper, lead, iron and silver (Salway 1993; Southern 2012). A phase of aggradation on Dartmoor dates between the 4th and 7th centuries AD, suggesting that either late Roman or early post-Roman mining was evident in the CWDML region (Thorndycraft *et al.* 2004). Although detailed Roman mining practices in Britain are not well recorded, the Romans are known to have employed enormous water wheels to drain the mines of Spain and Wales (Salway 1993; Southern 2012). This highlights that water was a problem for the miners, and also that surrounding aquatic environments would have been contaminated by the disposal of mining waste during this time.

Approximately 1,500 years later, during the medieval period (c. 410–1485 AD), mining for metals accelerated (Hatcher 1973; Lewis 1965). The principle method of extracting tin was through 'tin streaming,' where tin ore that has been exposed by weathering is washed by artificially flooding water into a valley, separating the tin ore

(cassiterite) from waste material (Lewis 1965; Gerard 2000). Smaller streams with steeper gradients were mined more often than larger rivers (Ehlen *et al.* 2005). Mine workings along larger streams however, could extend as far as 100m or more to either side of the channel (Newman 1998). Streaming usually moved upstream with miners working from the mouth of the stream up to its headwaters (Newman 1998). Tin streaming not only contaminated the water with tin and a variety of other toxic metals from mine-waste tailings but also released large amounts of sediment that was washed downstream (Kondolf & Piegay 2003; Pirrie *et al.* 2002; 2003). Salmonids are known to be extremely sensitive to high sediment load (see review by Kemp *et al.* 2011). Sediment can smother gravel redds, reducing the growth and survival of eggs and alveins (Resier & White 1998; Walling *et al.* 2003). It is also known to negatively affect adult physiology and behavior; it can cause gill damage (Herbert & Merkens 1961) and affect swimming performance, respiration (Berli *et al.* 2014) and importantly, growth and survival (Suttle *et al.* 2004).

Following exhaustion of the most obvious deposits, later exploitation required deeper opencast and underground shaft mining, which became a significant practice by around 1500 AD (Homer *et al.* 1991; Ehlen *et al.* 2005). The deeper the mines, the more they flooded, and the more water had to be pumped out (Leadbeater 2009). Miners were permitted to mine virtually anywhere and to divert watercourses (Homer *et al.* 1991). In 1314, Devon miners were said to be laying waste to over 300 acres a year and in 1361, there were complaints about water diversions and the vast amount of water being used for mining practices (Homer *et al.* 1991). Geochemical dating of sediment aggradations and deposits shows a clear peak in tin and sediment lode from the 13th century (Pirrie *et al.* 2002; Thorndycraft *et al.* 2004), suggesting that the mining practices of this time had a profound effect on the surrounding environment. It should be noted that, although tin was often the main targeted metal of mining, the metal itself in its inorganic form is not considered to have a significant toxic effect on aquatic organisms (Eisler 1989). However, the mining of tin would have mobilised the release of other toxic metals into the environment.

Table 1. Estimated total mineral and metal production from southwest England during the 19th Century. Adapted from LeBoutillier (2002) with information obtained from Dines (1956); Alderton (1993) and South Crofty PLC (1998-1999).

Metal	Approximate Tonnes
Tin (Sn) metal	2,770,000
Copper (Cu) metal	2,000,000
Iron (Fe) ore	2,000,000
Lead (Pb) metal	250,000
Arsenic (As) as arsenic oxide	250,000
Manganese (Mn) ores	100,000
Zinc (Zn) metal	70,000
Silver (Ag) ore	2,000
Silver (Ag) metal (from sulphide ore)	250

1.2.3 INDUSTRIAL MINING

During the height of the mining boom in the 19TH Century, over 50,000 miners worked more than 1,000 mines in southwest England (Dines 1956; Leifchild 1968). The region became the nation's powerhouse, supplying two thirds of the world's copper and half the world's tin. The peak of copper export was reached by the 1860s, when approximately 15,500 tons of metal were produced annually (Dines 1956), and tin production peaked in 1870 with over 10,000 tons of metal produced annually (Burt *et al.* 1987). Between the late 1800s and early 1900s, the annual output of arsenic was between 4,000 and 8,000 tons. Lead was produced in nearly all of the mining districts, with as many as 400 mines active at any one time (Dines 1956). Total estimated yields of metals from this period are detailed in Table 1.

Water remained the major preoccupation of mining engineers during this period. For example, in 1846 a catastrophic flooding of the shafts of East Wheal Rose mine near Newquay drowned 39 miners (Gamble 2011). Dozens of mine pump inventions were patented and huge beam engines were used to pump water out of the mines (Barton 1969; Navolari, 2004). At Godolphin Mine (surrounding the River Hayle – see section

1.2.4), five steam pump engines and two water pumps were required to keep water out of the mining shafts (Harris 1945). The large majority of floodwater removed from mines was simply released back into rivers (Best 1999). This highlights the colossal scale of mining activity that occurred over thousands of years in the southwest of England. The effects of such practices would have been catastrophic for the surrounding ecosystems and, in particular, on aquatic environments.

1.2.4 CONTEMPORARY CONTAMINATION AND STUDY RIVERS

For the majority of rivers affected by historical mining activity, approximately 90% of metal contaminants are associated with sediment, transported downstream by natural fluvial processes and in some cases, resulting in contamination problems tens of kilometres downstream from the mine site (Environment Agency 2013). It should be noted that although the rivers affected by mining are broadly referred to as 'metal-impacted' or 'metal-polluted', all contaminated rivers are not equal. There exists not only variance in the amount of metal contamination between rivers but also a complex of contrasting metal types within each metal-impacted river, which form 'cocktails' of different metals. The primary metal-impacted study rivers assessed in this thesis, the River Hayle and the Red River, are both polluted by high concentrations of metals in comparison to clean rivers. However, when comparing these two rivers, the Red River is characterised by contamination with a high amount of arsenic (Camm *et al.* 2004), while the River Hayle has higher concentrations of zinc and nickel (Brown 1977; Minghetti *et al.* 2014; Figure 6). It is important to appreciate the effects that this differential contamination of rivers would have had, and continue to have, on the native populations of resident trout, both in terms of tolerance to unique metal mixtures and in the relative decimation of ecosystems within each river.

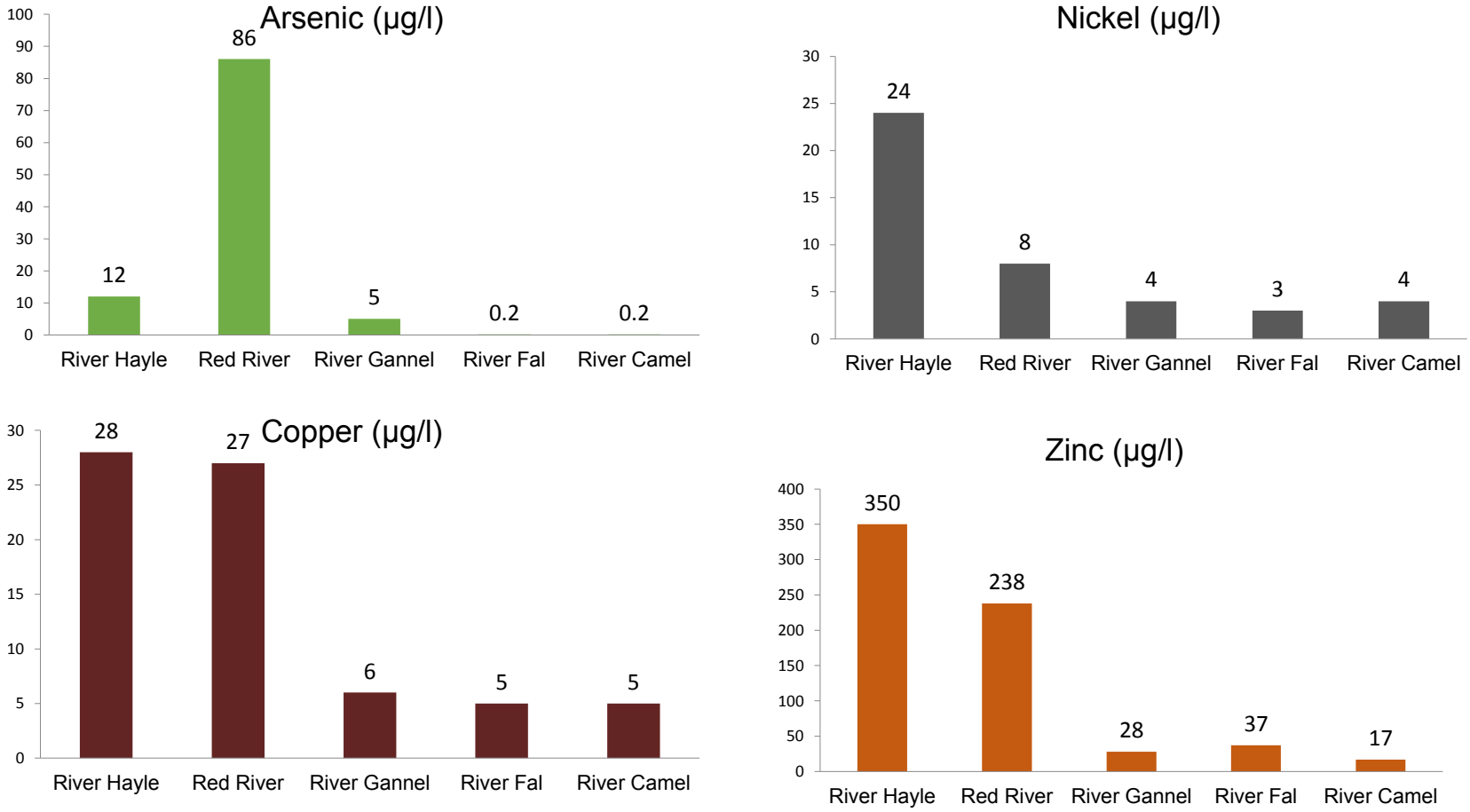


Figure 6. Metal contamination data for two metal-impacted rivers and three comparison clean rivers in southwest England. Data obtained by request from the Environment Agency.

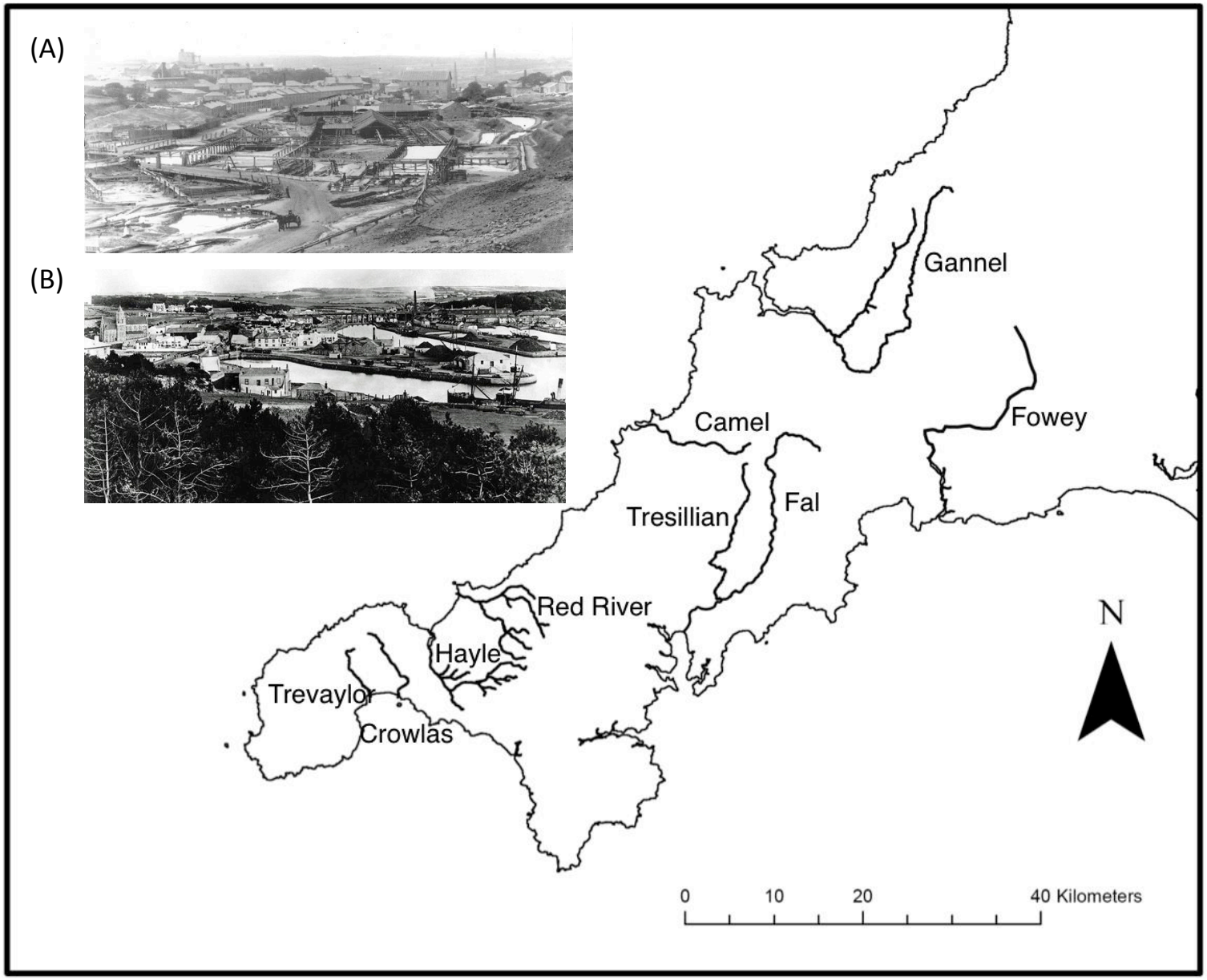


Figure 7. Map and images of the main metal-impacted study rivers: the Red River and the River Hayle. Pictures depict: (A) The Red River Valley in the 19th Century (from Buckley 2012); and (B) The River Hayle Estuary in the 19th Century (from Hayle Harbour Authority Operations Ltd)

The Red River (Cornish: Koner) is aptly named due to vibrant ochre colour of its water, a result of historical tin mining (Banks *et al.* 1997). Flowing through northwest Cornwall, the river is relatively small (~13 km long), yet the surrounding area has been extensively mined. Its catchment area covers three once very active mining districts (Camborne, Tuckingmill and Pool) and includes some of the longest-lived mines of the CWDML region (Dines, 1956). The Red River has suffered mining pollution both historically and recently (restorerivers.eu; accessed Nov. 2015). The entire catchment was extensively tin streamed and its water used for both mineral separation and as a source of power (Mitchell & Moore 2009). More recently, the lower catchment of the river received discharges from South Crofty mine between 1980 and 1988 (restorerivers.eu; accessed Nov. 2015). Tin-mine tailings in the Red River valley contain high concentrations of tin, copper, zinc, iron, manganese, arsenic and tungsten (Yim 1981).

The River Hayle (Cornish: Heyl), flows through the St Erth, Gwithian Tregonning, Godolphin and Gwinear mining districts. The river is only ~11 km long, yet dozens of mines existed across the catchment. Several of the mines within the area were substantial producers but many other smaller mines are also documented within this area (cornish-mining.org.uk; accessed Nov. 2015). The River Hayle catchment was first mined in the Bronze Age (Pirrie *et al.* 1999), and it was also tin streamed in the medieval period (Gamble 2011). Due to the 19th Century mining boom, the River Hayle today is characterised by a complex of different concentrations of metals: a moderately contaminated downstream region, an extremely contaminated middle section (at Godolphin) and less contaminated upper reaches (Figure 8). Mines surrounding the downstream portion of the river primarily produced copper: the Alfred Consols producing 160,000 tons in its lifetime (Dines 1956). The middle portion of the river was surrounded by the Godolphin mines; the area has been declared 'one of the best speculations for tin that can be found in any part of the Country' (Leifchild 1968). Contemporary concentrations are so elevated in the river area surrounding Godolphin that there is a ~3.5km dead zone of macroinvertebrate life (Brown 1977; Durrant *et al.* 2011). Present-day metal contamination in the river remains exceptionally high, especially in respect to concentrations of tin, arsenic, copper, lead, tungsten and zinc (Yim 1976; Turner 2000).

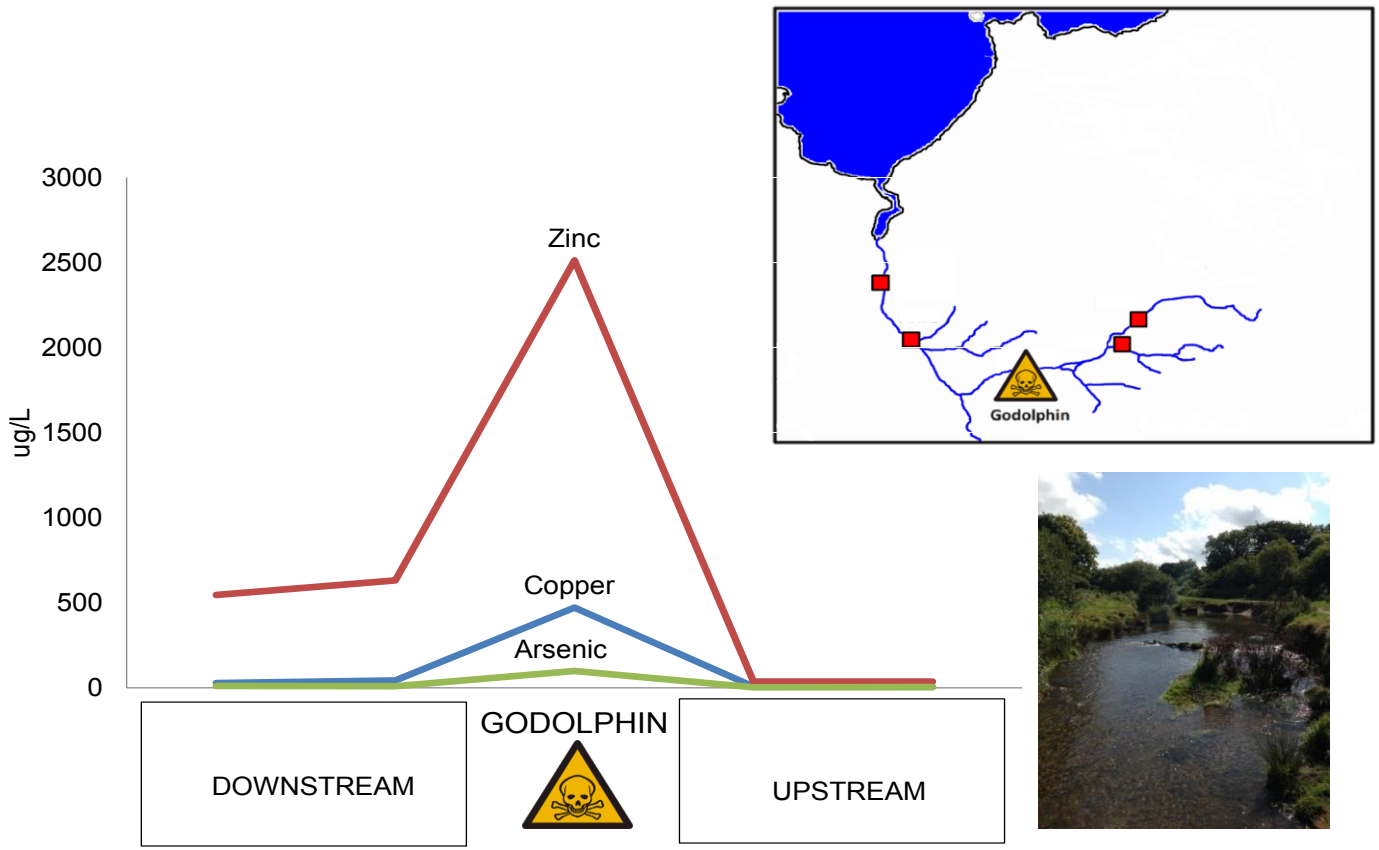


Figure 8. Patterns of metal contamination within the River Hayle. Concentrations of arsenic (As), copper (Cu) and Zinc (Zn) are markedly higher at Godolphin in the middle of the river. There are moderately high metal concentrations downstream of Godolphin and lower concentrations upstream of Godolphin.

Given salmonids' sensitivity to the environmental changes associated with this large-scale mining activity, it is likely that during intense periods of mining, local populations of brown trout either survived as intra-river meta-populations in pockets of unaffected stretches of river or, alternatively, became locally extirpated from entire river systems. Conceivably, this may have left a genetic signature in the make-up of these populations in the form of population bottlenecks, reduced genetic diversity and stochastic in-river population structure. Furthermore, it is perceived that recolonisation of the affected rivers may have occurred through local inter-river straying of sea trout, or from intra-river peri-refugial populations. We cannot ascertain for certain by which processes trout re-colonised these affected areas but, by exploring the patterns of genetic structuring and diversity across these populations, we can begin to build a picture of the effects this environmental stress would have had on the contemporary biology of trout that now reside in these rivers.

1.3 GENETIC ADAPTATION AND PHENOTYPIC PLASTICITY

Genetic local adaptation is defined as evolutionary patterns and processes acting on local populations of the same species that are potentially connected by dispersal and gene flow (Kawecki & Ebert 2004). Genetic local adaptation is essentially the playoff between the diversifying force of natural selection and the homogenizing forces of gene flow. In terms of natural selection, theory states that because of their effects on fitness, altered conditions arising from changing environments can be a powerful force in adjusting the mean of a trait by changing allele frequencies under directional selection (Hoffman & Hercus 2000). Such directional selection leads to a change in the genetic composition of the organisms in question and in turn affects the overall genetic architecture of the local population.

An alternative to the processes of genetic adaptation in conferring the ability to survive in stressful or challenging environments is phenotypic plasticity, which can be defined as the capacity of a single genotype to display flexible phenotypes in variable environments (Whitman & Agrawal 2009). In contrast to genetic adaptation, when a trait is plastic it can be adapted to many environments (Auld *et al.* 2009), and an individual can adjust its morphology, physiology, behaviour, or life history in response to changing environmental conditions (Whitman & Agrawal 2009). In addition, when phenotypes are plastic, individuals with identical genotypes may have different phenotypic responses to living in different environments (Freeman & Herron 2004). It is important to note that phenotypic plasticity and genetic local adaptation are not mutually exclusive and an adaptive strategy may involve both processes. This thesis aims to address the relative significance of both processes in relation to metal-tolerance in brown trout.

1.3.1 A HISTORY OF MEASURING GENETIC ADAPTATION AND PHENOTYPIC PLASTICITY

Classical tests to quantify the relative contribution of these processes have involved reciprocal transplant experiments, where individuals from a particular site are moved or reared in another site of interest. Traits are measured in individuals and, if they confer higher fitness in their 'home' site compared to the transplanted site, genetic adaptation and/or phenotypic plasticity are taken as having been evidenced.

Historically, this was the primary method for examining these processes, but difficulties in studying mobile species mean the method is generally more applicable in plants and other sessile species, e.g. two 30-year transplants of arctic plant species (Bennington *et al.* 2012). Despite this, the method is still used to study more complex experimental situations; for example, in fish (Plath *et al.* 2013; Defo *et al.* 2015) and terrestrial vertebrates (e.g. Ergon *et al.* 2001), although primarily in the form of semi-natural enclosures (Olsson & Shine 2002; Oksanen *et al.* 2012).

The field of quantitative genetics is based on the use of statistical tools to analyse the inheritance of phenotypic characters as well as their evolutionary response to selection (Pearson 1898; Weldon 1895; Orr 2005). This is a biometric or statistical approach in which phenotypic differences are measured as a proxy for the effects of the environment on background genetic variation and plasticity. The classical quantitative genetic model is concerned with measuring the phenotypic trait value of an individual in a population, and can be deconstructed into the heritable effect of genotype and the environment, such that Phenotype = Genotype + Environment. Each character is considered as a metric feature, under genetic control by several genes, all making an autonomous contribution to the overall phenotypic value of the character (van Straalen & Timmermans 2002).

Testing for such interactions between phenotype, environment and genotype often involves the use of Quantitaive Trait Loci (QTLs) and experimental crosses to associate markers between the phenotype of parents and offspring (QTL mapping). Location of QTLs contributing to complex traits is achieved through genetic crosses that result in gametic phase disequilibrium between molecular markers and QTLs with higher values of a particular trait statistically associated with a particular allele at a locus (Lynch & Walsh 1998). QTL quantification is a large field of research and an especially popular technique, employed for decades, in plant breeding (Collard & Mackill 2008), crop improvement (Collard *et al.* 2005) and livestock breeding programs (Dekkers 2004; Goddard & Hayes 2009). In the last few decades, QTLs have also been increasingly used to relate selection to molecular markers (marker-assisted selection or MAS; e.g. Lande & Thompson 1990) and associating divergences in population structure with these selective traits: Q_{ST} (Spitze 1993; Prout & Barker 1993); see McKay & Latta (2002) and Whitlock & Guillaume (2009) for a review.

The field of quantitative genetics has done much to quantify constraints on the rate and direction of responses to selection in natural populations. One well-known example is the classic study exploring the evolution of locally metal tolerant grasses (Antonovics *et al.* 1971), which showed that vegetation was considerably influenced by metal concentrations, and that genetic changes were responsible for colonisation of metal-contaminated soils (Macnair 1987; Wright *et al.* 2013). Klerks & Levinton (1989) demonstrated resistance to nickel and cadmium in a benthic oligochaete (*Limnodrilus hoffmeisteri*); exposure experiments showed statistically significant differences between the survival rates of worms occupying a contaminated and control site. By examining the resistance of second-generation worms in a selection experiment, it was suggested that resistance was likely genetic, rather than arising through the processes of phenotypic plasticity.

More recent quantitative genetic analyses couple exposure studies with other quantitative techniques, as well as investigating further factors, such as the cost of adaptation. Agra and colleagues (2010) compared two populations of the freshwater flea *Daphnia longispina* from distinct reservoirs, one neighbouring a disused mine and the other inhabiting a nearby clean reservoir. In addition to exposure experiments, they determined broad sense heritabilities of the tolerance traits and conducted feeding experiments, allowing an insight into the ecological cost of tolerance. Variability between the two sites in relation to acute copper tolerance was postulated to be related to the different selective pressures of the local habitats, leading to genetic adaptation and lower feeding rates of the tolerant population.

The use of quantitative genetics has been criticised for the way in which some of its underlying theoretical principles affect the modelling of phenotypic evolution (see Pigliucci & Schlichting 1997; Pigliucci 2006). Additionally, the genetic architecture of the phenotypic trait is itself treated as a 'black box', with no discernment of the number of genes that affect the trait, the individual effects of each gene, or the location of these genes within the genome (Dekkers & Hospital 2002). This criticism is perhaps more appropriate given the current availability and rapid development of molecular techniques, which have allowed researchers to lift the lid of the 'black box' (see sections 1.3.2 and 1.3.3).

It has been argued that the presence of different constraints and advantages in quantitative genetics and molecular markers means that they should be viewed as complementary rather than interchangeable (Klerks *et al.* 2011). Amongst six regional populations of white spruce (*Picea glauca*), variation for quantitative traits showed a strong adaptive response, whereas variability of a number of genetic markers followed neutral expectations, i.e. migration-drift equilibrium (Jaramillo-Correa *et al*. 2001). Indeed, quantitative genetics still plays a huge role in studies of genetic adaptation and phenotypic plasticity (e.g. Le Corre & Kremer 2012; Blanquart *et al.* 2012; Shaw & Etterson 2012) and the development of modern experimental approaches such as QTL analysis and experimental microbial evolution means that the field remains relevant to exploration of the effects of the environment in driving changes in local populations.

1.3.2 MOLECULAR MARKERS AND THEORY IN PHENOTYPIC PLASTICITY

1.3.2.1 PHENOTYPIC PLASTICITY

Phenotypic plasticity is defined as the response to changing environmental conditions of an individual during its lifetime; consequently, molecular markers for investigating the phenomenon primarily interrogate changes in gene expression. The process of phenotypic plasticity hinges on acclimation to an environment and thus involves the regulation and expression of key genes and pathways associated with tolerance to the environmental stressor. This is in line with one particular definition that states that plasticity is any form of environment-dependent gene expression; this can include gene regulatory processes that may have no gross phenotypic effects (DeWitt & Scheiner 2004).

1.3.2.2 TARGETED GENE EXPRESSION

Gene expression is defined as ‘the process by which information contained in genes is decoded to produce other molecules that determine the phenotypic traits of organisms’ (Hartl 2009). Methods quantifying gene expression can be broadly categorised into targeted and global approaches. The most widely used targeted approach is through RT-qPCR (Reverse-Transcriptase quantitative-Polymerase Chain Reaction). RT-qPCR involves the amplification of cDNA derived from RNA,

where the copy number of the fragment is concurrently quantified with high copy numbers representing upregulation of target RNA and lower numbers suggesting a downregulation. For example, the technique has been used to assess phenotypic responses to different photoperiods in plants. Three key regulatory genes were shown to control for the photoperiod of flowering time in different plant phenotypes in a response to long-day and short-day conditions in *Arabidopsis thaliana* and rice *Oryza sativa* (Hayama *et al.* 2003). Upregulation of two genes was shown to involve photoperiod responses in wild and domesticated pea (*Pisum sativum*), and an ortholog of one gene controlled photoperiod responses in lentil (*Lens culinaris*) to short-season environments (Weller *et al.* 2012).

1.3.2.3 MICROARRAYS

Targeted approaches such as qPCR require an *a priori* biological judgement of which genes are likely to be affected by the stressor. A global screening of gene expression profiles is only possible by examining a much larger and richer variety of transcripts. Microarrays were among the first techniques used for quantifying such changes. Microarray platforms contain thousands of genes from a single-focus species, allowing a concomitant analysis of expression in response to environmental and genetic factors.

Assessing stress responses in rainbow trout (*Oncorhynchus mykiss*), Krasnov *et al.* (2005) found 63% differentially expressed stress-responsive genes correlated with 48 genes of interest. A microarray developed for gene expression in response to changes in temperature, salinity and light regimes in the massive reef-building coral *Oribcella* (prev. *Montastrea*) showed the upregulation of unique genes related to each stressor, as well as general stress responses upregulated in all stress types (Edge *et al.* 2005). Turner *et al*. (2008) used an *Arabidopsis thaliana* microarray to uncover significant genetic differentiation between populations of the closely related species, *Arabidopsis lyrata,* growing on serpentine and granite soils. Analysis showed the differential expression of 586 genes, many of which related to local environmental variables such as low Ca:Mg ratios.

A common drawback of microarrays is that they are species-specific and can often only be used for the focal species the gene set was developed for, or for other phylogenetically close species within the same genera (Ouborg *et al.* 2010).

However, an Atlantic salmon (*Salmo salar*) microarray was used for characterising oestrogen-responsive genes in rainbow trout (*Oncorhynchus mykiss)*, identifying several well-known genes but also two alternative genes which were sensitive to oestrogenic exposure (Gunnarsson *et al.* 2007). A species-specific microarray developed for zebrafish (*Danio rerio*) was used to assess gene expression responses to heat in the phylogenetically distant coral reef fish *Pomacentrus moluccensis*, identifying 111 genes that were differentially expressed in response to elevated temperatures (Kassahn *et al.* 2007).

1.3.2.4 SAGE

A further method of obtaining global gene expression profiles is through SAGE (Serial Analysis of Gene Expression), whereby a series of cDNAs (tags) in unique positions are used to identify transcripts. The tags are linked together into longer molecules, which are subsequently cloned and sequenced; the number of times a tag is observed is proportional to the expression level of the corresponding gene. SAGE tags sequenced in *Arabidopsis* showed that several genes responsible for acclimation to cold environments were highly induced in the leaves but not in the pollen, suggesting poor stress tolerance to low temperatures in pollen (Lee & Lee 2003). In response to high and low calcium (Ca^{2+}) water in the green spotted pufferfish (*Tetraodon nigroviridis*), SAGE revealed genes responsible for the acclimation responses to environmental Ca^{2+} changes and osmoregulation (Pinto *et al.* 2010).

Despite these examples, global expression profiles for non-model species are now most easily achieved through RNA-seq (see section 1.3.4.1).

1.3.2.5 GO TERMS

A common technique used for associating changes in gene expression is the analysis of Gene Ontology (GO) Terms, which aims to find biological commonalities between genes. Given a set of genes of interest, GO analysis identifies clusters of genes related to specific biological processes. The aims of GO include developing a set of controlled, structured vocabularies (ontologies) to describe key domains of molecular biology and applying GO terms in the annotation of sequences, genes or gene products in biological databases. The project aims to provide a centralised

public resource allowing universal access to these ontologies (Gene Ontology Consortium 2004). Such databases allow for a holistic understanding of the biological processes common to a large number of biological systems and are extremely helpful in unpicking the underlying phenotypic responses and genetic methods of acclimation.

1.3.2.6 COSTS AND LIMITS OF PLASTICITY

The regulation of gene expression plays a pivotal role in complex phenotypes and how these phenotypes respond to challenges posed by the environment. However, despite its adaptive significance, plasticity is neither ubiquitous nor infinite (Snell-Rood *et al.* 2010). Moreover, the optimal response to a continuously changing environment is 'perfect plasticity' – whereby an organism can limitlessly shift its optimum depending on any given environment. The rarity of such phenomena suggests that plasticity must be constrained (Murren *et al.* 2015). Constraints to plasticity are generally grouped into two components – costs and limits – the former leading to reduced fitness when a trait is produced via plasticity rather than constitutively, and the latter leading to an inability to produce the optimal trait value – or 'perfect plasticity' (DeWitt *et al.* 1998). Both costs and limits involve evolutionary trade-offs.

Five categories of costs have been classified by Dewitt *et al.* (1998), and reiterated by Kleunen & Fischer (2005): maintenance costs, production costs, information-acquisition costs, developmental-instability costs and 'intrinsic' genetic costs (see review by Auld *et al.* 2009). Together, these costs are mirrored by evolutionary trade-offs, in terms of both energy and material expenses, resulting in non-optimal fitness in situations where long-term resilience is important to changing environments. Furthermore, it has been hypothesised that environmental stress might increase the magnitude of the costs due to resource limitation (Dorn *et al*. 2000; Huang *et al.* 2015).

In terms of limits, six main categories have been posited: information reliability limits, lag-time limits, developmental range limits, epiphenotype limits (DeWitt *et al.* 1998), plasticity-history limits (Van Kleunen & Fischer 2005) and ecological limits (Valladares *et al.* 2007), although debate surrounds the identification of some of these categories as representing limits alone (excluding costs). Furthermore, limits

are difficult to distinguish from costs due to the measurement of trait and fitness values (see Auld *et al.* 2010). Importantly, however, evolutionary trade-offs also exist in relation to limits, such that the ability to become plastic to a particular stressor may hinder the ability to become plastic to another.

Overall, the costs of plasticity may be an important underlying cause for a limit of plasticity (Auld *et al.* 2010) and costs and limits may make plasticity suboptimal compared with a compromised level of plasticity that is more economical (DeWitt & Scheiner 2004). Overall, the costs and limits to constitutively expressing genes for coping with environmental stress should be assessed if we are to understand the long-term stability of phenotypic plasticity.

1.3.3 MOLECULAR MARKERS AND THEORY IN GENETIC ADAPTATION

Advances in molecular biology have generated a plethora of molecular markers that can be used to investigate patterns of local genetic adaptation. Such genetic markers provide information on the allelic diversity at a given locus and, as such, offer a way to identify the relative effects of different genetic forces, such as selection, gene flow and genetic drift, when exploring genetic local adaptation.

1.3.3.1 ALLOZYMES

Allozymes were the first genetic markers widely employed to link molecular genetics and ecology but are nowadays rarely used. It has been posited (see Watt 1994) that allele variants of an enzyme (an allozyme) would reflect selection pressures from the environment. Early studies on *Drosophila* highlighted such variants (Lewontin & Hubby 1966; Hubby & Lewontin 1966), although it was unclear what balance of forces were responsible for the observed genetic variation. Allozymes have been used in microgeographical studies of local adaptive differentiation to environmental factors. The enzyme glycerate dehydrogenase (*Gly*) was used to investigate the genetic response to soil moisture patterns in the Colorado pinyon*, Pinus edulis* (Mitton & Duran 2004). *Gly* has subsequently been used to explore these patterns across a variety of spatial scales (see Jump & Penuelas 2005). Variation in phosphoglucose isomerase (*PGI*) showed defined shifts in allele frequencies selected for under different thermal regimes between lowland and alpine habitats in

populations of Mead's sulphur butterfly, *Colias meadii* (Watt *et al*. 2003; Wheat *et al*. 2005).

Allozyme analyses are criticised on the grounds that the level of differentiation among conspecifics is often low, and a substantial amount of genetic variation probably goes unnoticed; this is because of the redundancy of the genetic code, i.e. synonymous nucleotide substitutions, which may lead to homoplasy or convergent evolution when using such markers to explore phylogenetic/evolutionary relationships (e.g. Stevens & Godfrey 1992). Even when a non-synonymous substitution does cause an amino acid change, the new allele is often not apparent (Beebee & Rowe 2008). Furthermore, allozymes are functional proteins and therefore are not always selectively neutral. This can be an advantage or a disadvantage depending on the goal of the investigation, but non-neutral markers can be extremely useful if the aim is to examine evidence of adaptation (see section 1.3.3.3). Moreover, while both microsatellite and allozyme markers revealed a genetic distinction between Mediterranean and Atlantic populations of the ground beetle *Pogonus chalceus,* only an allozyme locus (*Idh1*) was variable between salt-marsh habitats (Dhuyvetter *et al*. 2004).

1.3.3.2 RFLPs, RAPDs and AFLPs

Molecular markers that focus on variation in DNA sequences, rather than at the protein level, now dominate because they are more direct markers, allowing quantification of the number of polymorphisms between different alleles (Schlötterer 2004). Dominant markers were the first of this type, although they are limited by the fact that they represent only a single dominant allele. RFLPs (Restriction Fragment Length Polymorphisms) have been used, for example, to study habitat preferences in the marine amphipod *Eogammarus confervicolus*, in which genetically-driven local preferences for wood habitats arose in less than 150 generations (Stanhope *et al.* 1993). RAPDs (Random Amplified Polymorphic DNA) are noteworthy, as they have also been used to study genetic adaptation, for example in contaminated sites in fish (Nadig *et al.* 1998), invertebrates (Ugolini *et al*. 2004) and plants (Mengoni *et al.* 2000). However, the popularity of RAPDs for studies of genetic adaptation has significantly decreased due to their lack of reproducibility (MacPherson *et al.* 1993).

Identification of loci linked to genes affected by genetic adaptation can be achieved by quantifying patterns of genetic differentiation between populations that occupy different environments. To do so with confidence however, requires the screening of a large number of loci to accurately estimate the expected level of genetic differentiation from other evolutionary forces and the proportion of loci linked to genes implicated in adaptation and reproductive isolation (Schlötterer 2003).

Such screening is now best achieved through whole-genome style analyses (see section 1.3.4). Before this, studies investigating adaptation in non-model organisms relied on the analysis of AFLPs (Amplified Fragment Length Polymorphisms) as the preferred method for obtaining a large number of loci (Campbell & Bernatchez 2004). AFLPs can be used as candidate markers for genomic regions involved in genetic adaptation (Wilding *et al.* 2001).

Evidence of genetic adaptation to high altitudes following the scanning of over 3,000 AFLPs in populations of the common vole (*Microtus arvalis*) showed that detection of AFLP markers under selection match the statistical power of that obtained by single nucleotide polymorphisms (SNPs; Fischer *et al.* 2011). AFLPs examined in populations of the blind mole-rat *Spalax galili* occupying sharply contrasting soil environments separated by just a few hundreds of metres revealed 211 soil-unique alleles (Polyakov *et al.* 2003).However, replacement of these dominant markers with more informative co-dominant markers means that their prevalence has significantly decreased in recent years.

1.3.3.3 MICROSATELLITES

The most commonly used neutral molecular markers for detecting genetic adaptation are microsatellites, although the ease of genotyping SNPs means that these markers may be used less in the future (Vignal *et al.* 2002; although see johnbhorne.wordpress.com: *Why microsatellites are still cool!*). The primary focus of genetic adaptation to local environments is on the contemporary genetic structuring of geographically close populations occupying different habitats and as such, requires DNA markers that have a high evolutionary rate (Wan *et al.* 2004). Microsatellites have an estimated average mutation rate of $\mu = 5 \times 10^{-4}$ (Whittaker *et al.* 2003), thereby generating the high levels of allelic diversity necessary for genetic studies of processes acting on recent and current time scales (Schlötterer, 2000).

Putatively neutral markers such as microsatellites can be very effective in detecting genetic adaptation, although by definition their neutrality means that these markers by themselves are not affected by selection (Klerks *et al.* 2011), unless they are physically linked (or close) to a part of the genome that is under selection (via genetic hitchhiking). Generally, variation in these markers only decreases when the population is isolated, or the size of the population dramatically declines and gene flow is disrupted.

The use of microsatellites in demonstrating patterns of genetic adaptation is widespread; a Web of Science search for 'microsatellites' and 'local adaptation' from to 2000-2016 produced 267 results (search conducted in October 2016). Strong genetic structuring is prevalent in salmonids, even in populations connected by gene flow, a pattern likely influenced by their natal homing instinct (Taylor 1991). Local adaptation has therefore been purported to be very high ('a paradigm in studies of salmonid fish' – Meier *et al.* 2011); a meta-analysis estimated the frequency of local adaptation in salmonid fishes to be ~55-70% (Fraser *et al.* 2011). A large body of work investigating local adaptation in salmonids using microsatellite markers supports this (but rarely shows a causal link).

For example, Hansen *et al.* (2002) analysed microsatellites in five anadromous populations of brown trout in Denmark and predicted the likelihood of genetic adaptation in the face of a lack of genetic drift (due to high effective population sizes - N_e) and high amounts of gene flow (high migration rates – m). They showed that populations still displayed significant population structure and moreover, that these patterns were temporally stable. Furthermore, the process of genetic drift has been shown to dominate the forces of genetic adaptation in contributing to the genetic structure in isolated populations (Campos *et al.* 2006). A population structure analysis of over 2,700 anadromous Atlantic salmon showed that hierarchical structuring of populations were driven by interactions between gene flow and adaptation to thermal regimes (Dionne *et al.* 2008). Early life-history traits measured in brown trout were more significantly divergent between populations than expected given the neutral forces of genetic drift and migration (Jensen *et al.* 2008).

1.3.3.4 MHC

An appropriate marker set will include loci that contain enough genetic variance in a particular trait to distinguish both between and within populations of the same species, whilst also possessing the ability to undergo substantial change in allele frequency with environmental change (Bickman *et al*. 2011). An alternative to such neutral markers is to determine changes in genes that are under selection, i.e. adaptive genes that are directly engaged in the response to environmental changes (Hoffman & Willi 2008). One such example is the major histocompatibility complex (MHC), which participates in immune response function and contains high genetic variability (Sommer *et al.* 2005). MHC genes have been utilised in many studies of genetic adaptation in fish (Eizaguirre *et al*. 2011; Ackerman *et al,* 2012; Evans *et al.* 2010). In particular, MHCs are thought to be linked to the presence of distinct parasite communities, which would in turn require contrasting local immunogenetic adaptation by the hosts (Eizaguirre *et al.* 2012).

1.3.3.5 TESTING FOR SELECTION

The primary method for detecting genetic adaptation using supposedly selectively neutral markers, such as microsatellites, is through the use of 'outlier tests', which use various statistical methods to determine if a locus is under selective forces in comparison to genetic drift and gene flow or is significantly out of Hardy-Weinberg Equilibrium (HWE). A large proportion of these outlier tests rely on statistical frameworks that detect acute F_{ST} outliers, such as distance-based methods (Beaumont & Nichols 1996; Beaumont and Balding 2004) and more complex implementations of these tests: Bayesian - Bayescan (Foll & Gaggiotti 2008); bisection approximation and contour smoothing - Lositan (Antao *et al.* 2008). Other selection tests involve coalescence simulations - DetSel (Viatlis *et al.* 2003), or ratio statistics – LnRH & LnRV (Schlötterer 2002). These outlier tests are extremely popular for identifying potential genetic adaptation in population genetics applications, although performance varies (Pérez-Figueroa *et al.* 2010; Lotterhos *et al.* 2014).

Outlier loci are purported to be under selection because the neutral marker is in linkage with a part of the genome under selection and is therefore an outlier due to

genetic hitchhiking (Maynard Smith & Haigh 1974) or may be (but rarely is) directly functional (e.g. the CAG segment in Huntington's disease). Significantly, however, for non-model species, it is often not possible to relate the locus directly to an adaptive function.

More recent outlier tests provide additional support for a possible adaptive function by correlating the outlier loci to environmental variables (genotype-by-environment interactions). Various methods have been developed for this, and such tests are often far more informative in linking markers that are under selection to their environmental regime than their classical antecedents. More recent applications of these methods include Random Forest algorithms, which have shown adaptation to habitat features in panmictic European eels (*Anguilla anguilla*; Pavey *et al.* 2015), and SNPs linked to run-timing in Chinook salmon (*Oncorhynchus tshawytscha).* Latent Fixed Effect Models (LFMMs) have demonstrated SNP-climate correlations in maritime pine (*Pinus pinaster;* Jaramillo-Correa *et al.* 2015) and specific loci related to pathogen load and salinity-driven selection in Atlantic salmon (Zueva *et al.* 2014).

1.3.4 GENOMICS FOR DETECTING PLASTIC AND GENETIC RESPONSES

The 'genomic revolution' that has occurred in the last two decades has dramatically increased the number of studies that are using a wide array of genomic techniques to answer critical questions in evolutionary biology. These advances in sequencing technology (i.e. high-throughput sequencing), along with the falling cost of such methods, have allowed a much more dense sampling of the genome, giving researchers the ability to observe the patterns of genetic variation that result from evolutionary processes acting across the entire genome (Allendorf *et al*. 2010; Stapley *et al*. 2010; Li *et al*. 2012).

This increased availability has allowed for the simultaneous analysis of huge numbers of loci and has given rise to an 'ecological genomics' (Savolainien *et al.* 2013) or 'adaptation genomics' (Stapely *et al.* 2010) approach to the study of genetic adaptation and phenotypic plasticity. This new approach facilitates improved characterisation of the evolutionary processes that influence variation across genomes and populations (Luikart *et al.* 2003).

1.3.4.1 RNA-SEQ & EPIGENETICS

RNA-Seq is a method for exploring phenotypic plasticity on a genome-wide scale. Complementary cDNA is transcribed from messenger RNA (mRNA) and is sequenced using high-throughput sequencing technology, resulting in thousands, to hundreds of thousands, of transcripts. These transcripts can be used to build a transcriptome *de novo*, and this method does therefore not require a reference genome. Differential expression profiles of replicates from experimental or natural manipulations can then be compared against this reference transcriptome, allowing the quantification of plastic responses of individuals to different environmental pressures.

This population transcriptomics approach is able to distinguish genes responsive to both environmental and genetic stresses, as well as determining candidate genes and gene pathways underlying phenotypic plasticity (Ouborg *et al*. 2010). The main advantage of RNA-Seq over traditional RNA techniques, such as microarrays and RT-qPCR, is that it is not limited to detecting transcripts that correspond only to existing genomic sequences (Wang *et al*. 2009), and genes of interest do not have to be identified *a priori,* which makes it a more attractive technique when studying plastic responses in non-model organisms and novel environmental scenarios.

The transcriptome of the red abalone (*Haliotis rufescens*), produced through RNA-seq, allowed exploration into the species' tolerance to global climate change, with identification of a large number of genes involved in response to temperature, pH and oxygen concentration (De Wit & Palumbi 2013). Under experimental conditions, the rainbowfish (*Melanotaenia duboulayi*) showed plastic regulatory responses to temperature stress, likely conferring the species' ability to occur in multiple ecoregions, even in the face of reduced dispersal (Smith *et al.* 2013). Both the number of genes involved in range-edge expansion, and also upregulation of genes related to heat stress, were higher in leading-edge populations of the damselfly *Ischnura elegans*, allowing the species to acclimate to new temperatures experienced during expansion (Lancaster *et al.* 2016).

More recently, the field of investigating plastic responses using genomics has taken a further leap with a growing appreciation of epigenetic mechanisms involved in generating flexible phenotypes (Bell & Spector 2011). Epigenetics can be defined as

cellular modifications and processes capable of inducing changes in genetic activity without altering the underlying code of the DNA sequence (Jablonka and Lamb 2002). Thus, epigenetics represents a change in phenotype without a change in genotype, and it has been proposed that environmentally-induced epigenetic variation mediates a large amount of observed phenotypic plasticity (Angers *et al.* 2010; Feil & Fraga 2012). Such processes are primarily investigated by interrogating mechanisms of DNA methylation, histone modification and non-coding RNAs (Blakely & Litt 2016).

Epigenetic variability measured in the asexual cyprinid fish *Chrosomus eos-neogaeus* showed remarkable patterns of environment-driven as well as stochastic epigenetic changes between populations inhabiting lakes and intermittent streams (Leung *et al.* 2016). Transgenerational plasticity was observed in populations of *Daphnia* in response to contrasting fish predators (Walsh *et al.* 2016). Analysis of adaptation to habitat type in two salt marsh perennial plant species showed no environmental correlation in AFLP genetic variation, but significant differences in methylation-sensitive AFLP markers (Foust *et al.* 2016).

The field of epigenetics represents a promising new area of research in unpicking the different molecular levels that produce phenotypic variation (Flatscher *et al.* 2012) and may also represent a significant temporal buffer for genetic adaptation (Putnam *et al.* 2016; O'Dea *et al.* 2016).

1.3.4.2 RADSEQ & DETECTING SELECTION

Restriction Site Associated DNA sequencing (RADseq) involves sequencing a reduced representation of the genome of several individuals, to create hundreds of thousands of loci anchored by a restriction-enzyme cut-site (Baird *et al.* 2008; Etter *et al.* 2011). The power of RADseq is that it can be applied in model (reference-genome-guided) and non-model organisms (*de novo* construction of sequences) alike. Originally developed for microarray platforms, the marriage of RADseq with high-throughput sequencing technology has given rise to a new and revolutionary system of sequencing-by-synthesis, which allows the screening of thousands of SNPs distributed across the genome. This method of detecting a high density of SNP markers has become extremely popular within the field of ecological and conservation genomics (see review by Narum *et al.* 2013).

For example, RAD-sequencing of the land snail, *Cepaea nemoralis*, revealed that phenotypic variation in shell colour (C) and banding (B) were associated with 44 anonymous RAD markers that were linked to the *C-B* supergene linkage group (Richards *et al.* 2013). Repeated adaptations of the groundsel, *Senecio lautus*, to different habitats across the Australian coastline were related to divergent SNPs linked to genes important for survival in coastal environments (Roda *et al.* 2013).

Despite the ease of sampling for SNPs using methods such as RADseq, detecting the underlying genetic mechanisms driving adaptations or phenotypes can still prove challenging. A prime demonstration of this is the numerous clinical applications that use genome-wide-association studies (GWAS) to detect patterns of gene-related diseases in humans (Manolio *et al.* 2009). For example, despite considerable resources, genetic correlations underlying the cause of coeliac disease remain largely cryptic (Dubois *et al.* 2010; Trynka *et al.* 2011).

A dominant strategy in detecting genetic adaptation in large genome-wide studies is examining large blocks of the genome that are in linkage disequilibrium (LD) and are therefore under selective processes. Extensive patterns of LD underlying adaptation to freshwater habitats in three-spine stickleback (*Gasterosteus aculeatus*) contribute to rapid colonisation of freshwater lakes (Hohenlohe *et al.* 2011; Catchen *et al. pers. comm.).* LD between immune system regulators and cancer regions have been implicated in both the susceptibility of Tasmanian devils (*Sarcophilus harrisii*) to Devil facial tumour disease (DFTD) and cancer risk in humans (Epstein *et al.* 2016).

However, despite these neat demonstrative patterns of strong selection and LD, the power and type of selection underlying most processes of genetic adaptation are still debated (Stephan 2016; Bernatchez 2016). These theories primarily concern the difference between adaptive traits, which are influenced by hard selection on just a few loci of large effect (and accompanying blocks of LD), or polygenic soft sweeps, whereby several loci of small effect act in synergy (Figure 9; Pritchard *et al.* 2010). Furthermore, the microevolutionary processes of selection acting on individual SNPs or loci are often not directly related to the primary processes underlying adaptation, but are instead due to the more infrequent mechanisms of macroevolution or punctuated equilibrium. For example, horizontal gene transfer (Hao & Golding 2006), large-scale genomic architectural traits (e.g. gene and genome duplication; Yeaman

et al. 2016) or copy number variants (CNVs; Sudmant *et al.* 2015) may be responsible. It is therefore important to appreciate that selection to local environments may act by one or several of these processes, and so exploring the patterns of all types of selection is paramount in understanding genetic local adaptation.

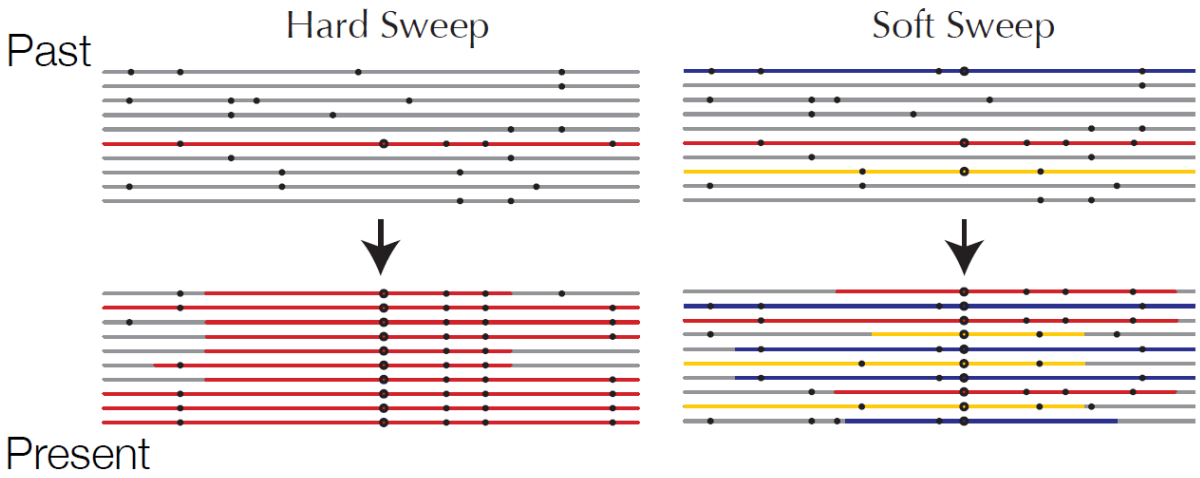


Figure 9. Sweeps of directional (or positive) selection across a genome, showing regions of the genome that also appear as under selection due to genetic hitch-hiking. Hard sweep (left) showing the selection of a single allele of large effect (in red). Soft sweep (right) showing the selection of three alleles of small effect (blue, red and yellow). Image kindly provided by Prof. Bill Cresko.

1.4 STATE OF THE ART

Given the theoretical basis and empirical examples of both genetic adaptation and plastic responses of species to different habitats and environmental stressors, investigating the underlying mechanisms of metal tolerance in trout in southwest England requires an assessment of the relative contribution of both processes: genetic adaptation and phenotypic plasticity. Furthermore, if the basis is adaptive, we need to assess how far down the evolutionary path towards recognition as a genetically distinct entity these metal-tolerant fish have advanced. Below, I review the plasticity of brown trout and their ability to diversify under different environmental conditions, together with a broader consideration of metal tolerance in fish in general. Finally, I summarise what research has previously been undertaken to understand the basis of metal-tolerance in brown trout in southwest England.

1.4.1 POLYTYPIC TROUT AND EVOLUTIONARY SIGNIFICANT UNITS (ESUs)

Throughout their range, brown trout display an astonishing mosaic of patterns of variation in phenotype and life history, which has led to major disagreements concerning population affinities and taxonomic boundaries in the assemblage (Avise 2000). Since Linnaeus' first description of *Salmo trutta* in 1758, over 50 distinct varieties (often suggested as new species) of brown trout have been named (Laikre *et al.* 1999; Behnke 2002). The species displays an extraordinary propensity to diversify into a range of morphological and ecological types (morphotypes and ecotypes respectively), lending it recognition as a 'successful polytypic species' (Elliot 1994).

The anadromous life history component represents perhaps the most extreme case of ecological divergence as, unlike salmon, only a proportion of resident fish smolt and go to sea (see section 1.1.3). This led biologists to originally define three distinct morphs: the resident trout (*Salmo trutta fario*); the sea trout (*Salmo trutta trutta*) and the lacustrine trout (*Salmo trutta lacustris*), which are now all recognised as ecological variants of the same species.

Another well-recognised example of a distinct trout morphotype is the ferox trout found in glacial lakes across England, primarily in Scotland and Ireland (fishbase.org). Once considered a separate species, *Salmo ferox* (Campbell 1979), the ferox morphotype, is defined by a switch to a piscivorous (rather than a planktivorous) feeding regime, a considerably larger body size (up to 100cm) and lengthened longevity – as long as 20 years (Finch 1998; Mangel 1996; Went 1979). Allozyme variation (Ferguson & Mason 1981), microsatellite variability (Prodöhl *et al.* 1992), mtDNA RFLPs (McVeigh *et al.* 1995) and allele frequencies of the *LDH-C1* locus (Hamilton *et al.* 1989; Ferguson & Taggart 1991) provide evidence of the genetic divergence of this morphotype and suggest that the ferox represents a genetically distinct and reproductively isolated variety of the brown trout species.

It is also notable that other forms of trout of are also recognised in the glacial lakes of Ireland: the gillaroo and sonaghen trout found in Lough Melvin, the latter classified by the IUCN as a new species in 2013 (*Salmo nigripinnis*), and the dollaghan trout (Lough Neagh, Northern Ireland). Similarly, the Ohrid trout, endemic to the Ohrid Lake in Albania and Macedonia (FYROM), which undergoes lacustrine migration, is also now recognised as a separate species by the IUCN (*Salmo letnica*), as is the marmoratus trout, identified by its distinctive marbled pattern (*Salmo marmoratus*), found in the rivers around the Adriatic sea (Crivelli 2006).

Given this remarkable ability of the brown trout species to diversify into morphotypes and ecotypes (some even going as a far as being reclassified as species in their own right), what is the basis for regarding these unique trout types as evolutionary distinct and, further, do they require specific conservation and management efforts?

Evolutionary Significant Units (ESUs) can be loosely defined as a group of organisms that have been isolated from other conspecific groups for a sufficient period of time to have undergone *meaningful* genetic divergence from these other groups (Ryder 1986; Moritz 1994). In terms of conservation, an 'ESU-like' unit is a population that merits separate management and has a high priority for conservation, associated with distinct population segments that should receive particular protection (Crandall *et al.* 2000). A review of ESU concepts by Fraser & Bernatchez (2001) argues that, despite long-standing debate and numerous conflicts amongst the community about what exactly comprises an ESU, the various ESU

definitions should be unified into an idea of *adaptive evolutionary conservation* (AEC). In particular, there is an emerging consensus that both adaptive divergence and historical isolation should be considered when designating such conservation priorities (Fraser & Bernatchez 2001).

In acknowledgement of the complex debate over definitions and semantics, within the context of this thesis I wish to emphasis the ESU synthesis presented by Waples (1991), who evaluated the ESU term in relation to the conservation of Pacific salmon (*Oncorhynchus* spp.). Waples' review touches on the importance of ecological or genetic diversity and the potential loss of diversity in respect of the species as a whole. In particular, he poses the questions: i) Is the population genetically distinct from other conspecific populations? ii) Does the population occupy unusual or distinctive habitat? and iii) Does the population show evidence of unusual or distinctive adaptation to its environment? It could be argued that the ESU concept (similar to the definition of a species) can be represented as a continuous line through time, and that these distinguishable morphotypes and ecotypes of brown trout can be placed at points along this line in terms of being recognised within our established definitions of distinct ESUs, sub-species or even as entirely new species.

1.4.2 METAL TOLERANCE IN FISH

Metal-tolerance in fish is apparent in several phylogenetically different species across the world. Fish populations occupying metal-impacted environments frequently display high concentrations of metals in various tissues (see Fatima *et al.* 2014), with different tissues conferring different responses to a wide range of metals. The gills are in intimate contact with the water and are the main point of entry for dissolved metals, and thus are a major target for metal toxicity (Olsson *et al.* 1998); the liver and kidney are the primary organs of storage, detoxification and excretion (Wood *et al*. 2012); the gut primarily handles dietary metal uptake (Lawrence & Hemingway 2003).

Assessing metal-tissue-burden data offers a physiological indication of the sites of toxicity, storage and routes of uptake (Langton & Spence 1994). For example, in a reciprocal transplant of yellow perch (*Perca flavescens*), kidney cadmium and copper concentrations of clean fish transported to a metal-contaminated lake increased with

exposure, but metal-tissue depuration did not occur in contaminated fish in the clean lake, suggesting that the contaminated population displayed limited plastic responses (Defo *et al.* 2015). Exposed individuals among fathead minnow (*Pimephales promelas*) showed accumulation of copper and cadmium in various target tissues but, in liver tissue, copper accumulation was significantly reduced by the presence of cadmium (Driessnack *et al.* 2016). In a meta-analysis on the effects of cadmium and copper in fish species, no studies were found that indicated adverse effects without increases in whole-body concentration of metals (Meador 2015).

As the gill is the primary organ involved in waterborne metal-uptake, a common detrimental effect of metals involves disturbances in ion regulation and osmoregulation, in particular of sodium (Na^+), chloride (Cl^-) and ammonia (NH_3) homeostasis (Grosell *et al.* 2002; Rogers *et al.* 2003; Zimmer *et al.* 2012). Osmoregulatory disruption by metals often involves inhibition of the enzyme transporter Na^+/K^+-ATPase (Heath 1995), as well as metals competitively inhibiting the transfer of essential ions across the gill (Wood 2012) and actively competing with uptake of essential metals through ionic mimicry (Ballatori 2002; Bridges & Zalups 2005).

Measurements of Na^+ and Cl^- flux (see Wood 1992) and quantifying Na^+/K^+-ATPase activity (see Kramer *et al.* 1986) provide sensitive methods of monitoring such disruptions. Reductions in Na^+/K^+-ATPase activity and losses of Na^+ and Cl^- are commonly observed as a result of metal toxicity (Pelgrom *et al.* 1995; Kulac *et al.* 2013; Silva & Martinez 2014; Wu *et al.* 2015). On the other hand, modifications of branchial mechanisms (McDonald & Wood 1993) have been demonstrated as a method of acclimation, or plasticity, to metal toxicity (Reid *et al.* 1991; Galvez & Wood 2002; Taylor *et al.* 2003).

In relation to gene expression, a common acclimation response to metals often involves the upregulation of important metal-binding and detoxification proteins such as metallothioneins (MTs) and glutathione (Chowdhury *et al.* 2005; Jozefczak *et al.* 2012). In particular, metallothioneins play an important role in the transport and storage of heavy metals (Hamilton & Mehrle 1986; Roesijadi 1992) and, as such, are a commonly used biomarker in studies of metal toxicity and pollution (Tom *et al.* 2004; Sarkar *et al.* 2006). Metallothioneins are high-molecular weight, cysteine-rich

proteins, which have the capacity to bind both physiological or 'essential' (e.g. copper, zinc) and xenobiotic or 'non-essential' (e.g. arsenic, cadmium) metals. As a result of the whole genome duplication event in salmonids, two isoforms of metallothionein are recognised: metallothionein A (MTA) and metallothionein B (MTB) (Bonham *et al.* 1987).

Gene expression of hepatic MT was upregulated in natural populations of mercury-tolerant fathead minnows (*Pimephales promelas*) and control individuals exposed to mercury-enriched water (Hamilton & Mehrle 1986). Copper-acclimated groups of the least killifish, *Heterandria formosa*, had significantly higher metallothionein protein concentrations compared to control cohorts (Adeyemi & Klerks 2013).

In terms of genetic responses, the study of metal-tolerance in populations of yellow perch (*Perca flavescens*) provides an ideal system for analysis. These populations exhibit an overall reduction in genetic diversity correlated with cadmium contamination (Bourret *et al.* 2008). Targeted SNP analysis between clean and metal-impacted populations identified 87 candidate coding SNPs, including outlier SNPs related to fast life-cycle completion (Bélanger-Deschênes *et al.* 2013). Moreover, these outlier SNPs were also correlated with cadmium concentration. A candidate-gene microarray showed 287 and 176 differentially expressed genes in response to different concentrations of nickel and cadmium respectively (Bougas *et al.* 2013).

1.4.3 METAL TOLERANCE IN TROUT IN SOUTHWEST ENGLAND

Different approaches to investigation of metal-tolerance in trout in southwest England have been considered above. Below, I provide a critical evaluation in order to assess what is currently known and therefore what further research is required. This will lead on to the main aims and objectives of this thesis.

Durrant and colleagues (2011) used complementary methods for exploring both the genetic adaptation and physiological plasticity of metal-tolerant trout in the River Hayle. This study aimed to investigate the extent to which the highly contaminated middle region of Godolphin was toxic to metal-naïve fish and also if it represented a chemical barrier to gene flow within the river.

This was quantified by assessing i) Na^+ influx rates of metal-naïve trout exposed to metal-contaminated water and ii) neutral genetic patterns using seven microsatellites. Results showed a significant reduction in Na^+ influx rates in metal-naïve trout exposed to metal concentrations typical of Godolphin water, showing that metal-naïve fish experience disruptions in osmoregulation. Unfortunately, Na^+ influx rates were not measured in metal-tolerant Hayle trout, so a direct comparison could not be made.

Analysis of the patterns of genetic structure showed that Godolphin did not act as a chemical barrier to movement of fish within the river and Hayle trout did not show evidence of reduced genetic diversity or population bottlenecks. Within-river genetic structure did not correlate with known levels of metal pollution and divergence between the differentially polluted regions of the river was low. The study therefore failed to find a mechanism responsible for tolerance to metals but proposed mechanisms involving either plasticity or an adaptive evolutionary response.

To investigate phenotypic plasticity between metal-tolerant fish from the Hayle and trout from a clean river (the River Teign), Uren-Webster *et al.* (2013) assessed metal-tissue burden to evaluate the role of different tissues in storage and detoxification of metals, and also used RNA-seq to quantify differential gene expression within each of these tissues. Significantly elevated concentrations of metals were found in the gill, liver and kidney of Hayle fish. Uptake, storage and metabolism of iron, cadmium and arsenic were strongly correlated. RNA-seq analysis revealed 998 differentially expressed transcripts, with the greatest number occurring in the kidney. Upregulated genes were linked to metal- and ion-homeostasis, and also to metabolic processes. Interestingly, very few upregulated genes were associated with oxidative stress or cellular damage and there was a downregulation of genes involved in an immune system response. Of the two metallothionein isoforms, MTB was one of the most strongly upregulated genes, whereas MTA displayed only low levels of expression.

This study was the first to demonstrate the role of plastic responses in the form of gene expression differences in driving the metal-tolerance observed in trout in southwest England. However, further investigation requires a more holistic analysis of metal-tissue burden and changes in gene expression of metal-tolerant trout in

control (clean) freshwater conditions and, similarly, measurement of the responses of clean-river fish to metal-contaminated water.

In order to assess the use of cultured gill tissue subjected to metal-contaminated water, Minghetti *et al.* (2014) employed a Fish Gill Cell System (FIGCS) to determine branchial epithelial response to aquatic toxicants. Cultured gill obtained from farmed rainbow trout (*Oncorhynchus mykiss*) were exposed to river-water from the Hayle and the transepithelial electrical resistance (TEER; a measurement of the health of a monolayer of cells) was quantified, as well as expression of both metallothionein isoforms (MTA and MTB) as measured by RT-qPCR.

Results showed a significant increase in the % TEER of FIGCS exposed to Hayle water from the upstream portion of the Hayle compared to all other treatments. MTA and MTB expression was significantly increased in FIGCS exposed to water from Godolphin in comparison to clean soft water. Furthermore, FIGCS exposed to Godolphin water showed significantly greater MT expression compared even to FIGCS exposed to water from the upstream region of the Hayle (where metal levels are lower). MTA mRNA expression was elevated in FIGCS exposed to water from all sites downstream of Godolphin, whereas MTB mRNA expression was only elevated in FIGCS exposed to water at Godolphin.

1.4.4 CONCLUSIONS

As pointed out by Merilä and Hendry (2014), separating the relative influences of genetic adaptation and plasticity underlying phenotypic change is extremely challenging. Nonetheless, several studies have employed techniques to investigate the role of both processes. Investigation of temperature tolerance in the killifish (*Fundulus heteroclitus*), showed that patterns of gene expression and non-neutral patterns of variation act on different sets of genes, yet the subset of genes that demonstrated both processes exhibited a counter-gradient of expression (Dayan *et al.* 2015). In guppy species of the genus *Poecilia*, inhabiting sulfidic and nonsulfidic interconnected habitats, microsatellites revealed strong population structure and restricted gene flow ('isolation-by-adaptation', or IBA) and a reciprocal transplant confirmed that resident fish in each habitat had higher probabilities of survival compared to immigrant fish (Plath *et al.* 2013). However, in a reciprocal transplant of

2,500 brown trout embryos, where there was no *a priori* knowledge of environmental factors driving divergent selection, early life history traits showed significant population differentiation and reduced genetic variation, but no evidence for genetic adaptation was found (Stelkens *et al.* 2012).

1.5 AIMS AND OBJECTIVES

The overall objective of this thesis is to adopt a multifaceted approach in assessing the patterns and processes involved in the basis of metal tolerance in brown trout populations in southwest England. I seek to establish the relative contribution of both genetic adaptation and phenotypic plasticity in driving both individual and population-level differences in trout associated with metal tolerance. Importantly, through analysis of the data chapters presented here, I assess the following questions and hypotheses:

Genetic adaptation and phenotypic plasticity

What is the underlying basis of metal tolerance?

H_1 Genetic adaptation is the basis of metal tolerance

H_2 Phenotypic plasticity is the basis of metal tolerance

H_3 Metal tolerance is the result of both genetic adaptation and phenotypic plasticity

Evolutionary significant units

How far down the evolutionary road of becoming an ESU are metal-tolerant trout populations?

H_1 Metal-tolerant trout populations are genetically distinct and represent distinct ESUs

H_2 Metal-tolerant trout populations are not genetically distinct and do not represent distinct ESUs

Costs of metal-tolerance

Has tolerance to metals forced these trout populations down an 'evolutionary cul-de-sac'?

H_1 Metal-tolerant trout have a limited genetic repertoire, which will affect their ability to adapt to future change in their environments

H_2 Metal-tolerant trout do not have a limited genetic repertoire and their ability to adapt to future change in their environments is not comprised

Conservation and management

What are the applied conservation and management recommendations for metal-tolerant trout populations?

H_1 Metal-tolerant trout do not represent unique populations and do not require tailored management and conservation actions

H_2 Metal-tolerant trout populations should be granted protection as unique populations representing ecological diversity within the trout species

It may not be possible to answer these questions definitively in this thesis, but I hope that an assessment of the data presented here will contribute to understanding of the ability of natural fish populations to evolve in human-altered landscapes. In this thesis, investigation into the patterns and processes underlying metal tolerance in trout will be addressed in in the format of five individual research-style papers:

1.5.1 Research Paper 1: Patterns of metal tolerance using microsatellites

Josephine R. Paris, R. Andrew King and Jamie R. Stevens (2015) Human mining activity across the ages determines the genetic structure of modern brown trout (*Salmo trutta* L.) populations. *Evolutionary Applications,* **8**(6): 573-585.

I use twenty-three microsatellite loci to investigate population genetic differences between metal-impacted and clean-river trout populations across southwest England. I use both neutral microsatellite markers (twenty-one loci) and microsatellite markers that have been linked to immune-system (MHC) genes (two loci). Fifteen populations of trout (nine metal-impacted, six clean control populations), totalling 700 individuals, are genotyped at 25 microsatellite loci. I conduct approximate Bayesian computation (ABC) in order to identify if genetic differences between populations coincide with known historical periods of increased mining activity.

1.5.2 Research Paper 2: Method development for RADseq analysis

Josephine R. Paris, Jamie R. Stevens and Julian M. Catchen (2017) Lost in parameter space: A road map for Stacks. *Methods in Ecology and Evolution* DOI: 10.1111/2041-210X.12775

To critically assess the role of selection in driving metal-tolerance in trout populations, it is essential to assemble biologically reliable loci from large, genome-wide datasets. I develop a method of *de novo* parameter optimisation for assembling RAD loci using the software *Stacks*. I also demonstrate a method of integrating genome-alignment information by using trout RAD data aligned to the reference genome of the Atlantic salmon (*Salmo salar*). This bolsters the amount of biological information that can be obtained by assessing potential patterns of selection through the use of genome scans. In order to confirm that the methods are robust, I also analyse two other RADseq datasets, obtained from the king penguin (*Aptenodytes patagonicus)* and the red earthworm (*Lumbricus rubellus*). I use the methods developed here to create a biologically robust and informative RADseq dataset for the following data chapter.

1.5.3 Research Paper 3: Patterns and processes of genetic adaptation using RADseq

Josephine R. Paris, Karen Moore, R. Andrew King, Julian M. Catchen and Jamie R. Stevens. Rapid and repeated local adaptation of brown trout populations to unique cocktails of metal contaminants *Manuscript in preparation.*

This is the first chapter that makes the shift from assessing general patterns of genetic diversity into uncovering the genetic processes underlying metal-tolerance in brown trout. I conduct RADseq analysis on 100 trout from a selection of both clean and metal-impacted rivers across Southwest England. Importantly, I assess the role of loci under divergent selection in driving population-level divergence and consider the gene ontology of these loci in order to relate them to biologically functional information. I use Latent Factor Mixed Models (LFMM) to explore the role of polygenic selection sweeps in metal-tolerance that, importantly, are statistically linked to metal-contamination profiles specific to each river.

1.5.4 Research Paper 4: Processes of phenotypic plasticity in physiological metal handling

Josephine R. Paris, Mauricio M. Urbina, R. Andrew King, Darren Rowe, Eduarda M. Santos, Nic R. Bury, Rod W. Wilson and Jamie R. Stevens. Brown trout display multi-faceted physiological mechanisms to cope with metal toxicity in a chronically metal-impacted river. *Manuscript in preparation*.

This chapter explores the potential physiological processes allowing trout to cope with metal-enriched waters. Importantly, measuring these physiological mechanisms lends insight into the relative roles of phenotypic plasticity and genetic adaptation. We conducted a laboratory experimental where 200 trout (100 from the River Hayle and 100 from a clean reference river, the River Fowey) were exposed to an acute metal-mixture. Throughout the 96-hour exposure, we sampled fish to assess a series of physiological time points, related to in-river physiology, depuration and re-exposure to metals. I use several parameters commonly employed in fish physiology and ecotoxicology in order to assess the phenotypic processes underlying metal-tolerance.

1.5.5 Research Paper 5: Processes of phenotypic plasticity using RNA-seq

Josephine R. Paris, Sophie Shaw, Darren Rowe, R. Andrew King, Jamie R. Stevens and Eduarda M. Santos. Natural populations of chronically exposed metal-tolerant trout display minimal hepatic transcriptional responses to an acute metal exposure. *Manuscript in preparation.*

In this chapter, I assess the role of phenotypic plasticity in terms of hepatic gene expression in conferring the ability of metal-tolerant fish to cope in their native environment. The liver was chosen as a vital organ involved in metal homeostasis and detoxification in fish. The samples for this analysis were obtained from 24 fish at the end of the 96-hour metal exposure as outlined in Research Paper 4. Fish livers were sampled from four groups: Fowey-control; Fowey-exposed; Hayle-control and Hayle-exposed. I analyse pairwise comparisons of gene expression between these

groups and relate the patterns of over- and under-expression of genes related to the metal exposure to biological processes using Gene Ontology (GO) Term analyses.

REFERENCES

Aarestrup, K., & Jepsen, N. (1998). Spawning migration of sea trout (*Salmo trutta* L.) in a Danish river. *Hydrobiologia*, **371**, 275–281.

Ackerman, M.W., Templin, W.D., Seeb, J.E. & Seeb, L.W., (2013). Landscape heterogeneity and local adaptation define the spatial genetic structure of Pacific salmon in a pristine environment. *Conservation Genetics*, 14(2), 483-498.

Adeyemi, J. A., & Klerks, P. L. (2013). Occurrence of copper acclimation in the least killifish *Heterandria formosa*, and associated biochemical and physiological mechanisms. *Aquatic Toxicology*, **130**, 51–57.

Agra, A. R., Guilhermino, L., Soares, A.M.V.M. & Barata, C., 2010. Genetic costs of tolerance to metals in *Daphnia longispina* populations historically exposed to a copper mine drainage. *Environmental Toxicology and Chemistry*, **29**(4): 939–946.

Alderton, D.H.M. (1993). Mineralization associated with the Cornubian granite batholith. In: Pattrick, R.A.D. & Polya, D.A. (eds) *Mineralization in the British Isles*, London: Chapman & Hall, pp. 270-354.

Allendorf, F.W. & G.H. Thorgaard, (1984). Tetraploidy and the evolution of Salmonid fishes. In: B.J. Turner (ed.) *Evolutionary genetics of fishes*. New York: Plenum Publishing Corporation, pp. 1-53.

Allendorf, F.W.*,* Hohenlohe, P.A.*,* Luikart, G., *(*2010*).* Genomics and the future of conservation genetics*. Nature Reviews Genetics,* **11***:* 697–709*.*

Almodovar, A., & Nicola, G. G. (2004). Angling impact on conservation of Spanish stream-dwelling brown trout *Salmo trutta*. *Fisheries Management and Ecology*, **11**(3–4), 173–182.

Almodóvar, A., Nicola, G. G., Ayllón, D., & Elvira, B. (2012). Global warming threatens the persistence of Mediterranean brown trout. *Global Change Biology*, **18**(5), 1549–1560.

Almodovar, A., Nicola, G. G., Elvira, B., & Garcia-marin, J. L. (2006). Introgression variability among Iberian brown trout Evolutionary Significant Units: the influence of

local management and environmental features. *Freshwater Biology*, **51**(6), 1175–1187.

Angers, B., Castonguay, E., & Massicotte, R. (2010). Environmentally induced phenotypes and DNA methylation: how to deal with unpredictable conditions until the next generation and after. *Molecular Ecology*, **19**(7), 1283–1295.

Antao, T., Lopes, A., Lopes, R. J., Beja-Pereira, A. & Luikart, G. (2008). LOSITAN: A workbench to detect molecular adaptation based on a FST-outlier method. *BMC Bioinformatics*, **9**(1), 323.

Antonovics, J., Bradshaw, A. D., & Turner, R. G. (1971). Heavy metal tolerance in plants. *Advances in Ecological Research*, **7**, 1–85.

Auld, J.R., Agrawal, A.A., & Relyea, R.A., (2009). Re-evaluating the costs and limits of adaptive phenotypic plasticity. *Proc. R. Soc. B*. **277**(1681): 503-511.

Avise, J.C. (2000) *Phylogeography: The history and formation of species.* Cambridge, MA and London, UK: Harvard University Press.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E.U., Cresko, W.W., & Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, **3**(10), e3376.

Ballatori, N. (2002). Transport of toxic metals by molecular mimicry. *Environmental Health Perspectives*, **110**(5), 689–94.

Banks, D., Younger, P. L., Arnesen, R.-T., Iversen, E. R., & Banks, S. B. (1997). Mine-water chemistry: the good, the bad and the ugly. *Environmental Geology*, **32**(3), 157–174.

Barson, N. J., Aykanat, T., Hindar, K., Baranski, M., Bolstad, G. H., Fiske, P., Jacq, C., Jensen, A.J., Johnston, S.E., Karlsson, S., Kent, M., Moen, T., Niemelä, E., Nome, T., Næsje, T., Orell, P., Romakkaniemi, A., Sægrov, H., Urdal, K., Erkinaro, J., Lien, S., & Primmer, C.R. (2015). Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature*, **528**(7582), 405–408.

Barton, D.B., (1966). The Cornish beam engine: a survey of its history and development in the mines of Cornwall and Devon before 1800 to the present day, with

something of its use elsewhere in Britain and abroad. Truro, Cornwall: The Author, Frances St.

Beaumont, M. A., & Balding, D. J. (2004). Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**(4), 969–80.

Beaumont, M. A., & Nichols, R. A. (1996). Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London B: Biological Sciences*, **263**(1377).

Beebee, T.J.C. & Rowe, G., (2008*). An introduction to Molecular Ecology*. Oxford, New York: Oxford University Press.

Bélanger-Deschênes, S., Couture, P., Campbell, P. G. C., & Bernatchez, L. (2013). Evolutionary change driven by metal exposure as revealed by coding SNP genome scan in wild yellow perch (*Perca flavescens*). *Ecotoxicology*, **22**(5), 938–957.

Bell, J. T., & Spector, T. D. (2011). A twin approach to unravelling epigenetics. *Trends in Genetics*, **27**(3), 116–125.

Belpaeme, K., Delbeke, K., Zhu, L., & Kirsch-Volders, M. (1996). Cytogenetic studies of PCB77 on brown trout (*Salmo trutta fario*) using the micronucleus test and the alkaline comet assay. *Mutagenesis*, **11**(5), 485–492.

Bennington, C. C., Fetcher, N., Vavrek, M. C., Shaver, G. R., Cummings, K. J., & McGraw, J. B. (2012). Home site advantage in two long-lived arctic plant species: results from two 30-year reciprocal transplant studies. *Journal of Ecology*, **100**(4), 841–851.

Berg, O. K., & Berg, M. (1987). Migrations of sea trout, *Salmo trutta* L., from the Vardnes River in northern Norway. *Journal of Fish Biology*, **31**(1), 113–121.

Berli, B. I., Gilbert, M. J. H., Ralph, A. L., Tierney, K. B., & Burkhardt-Holm, P. (2014). Acute exposure to a common suspended sediment affects the swimming performance and physiology of juvenile salmonids. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, **176**, 1–10.

Bernatchez, L. (2001). The evolutionary history of brown trout (*Salmo trutta* L.) inferred from phylogeographic, nested clade, and mismatch analyses of mitochondrial DNA variation. *Evolution*, **55**(2), 351–379.

Bernatchez, L. (2016). On the maintenance of genetic variation and adaptation to environmental change: considerations from population genomics in fishes. *Journal of Fish Biology*, **89**: 2519–2556.

Berrebi, P. (2015). Three brown trout *Salmo trutta* lineages in Corsica described through allozyme variation. *Journal of Fish Biology*, **86**(1), 60–73.

Best, G.A., (1999). *Environmental Pollution Studies*. Liverpool: Liverpool University Press.

Bickham, J. W. (2011). The four cornerstones of Evolutionary Toxicology. *Ecotoxicology*, **20**(3), 497–502.

Birkeland, K. (1996). Consequences of premature return by sea trout (*Salmo trutta*) infested with the salmon louse (*Lepeophtheirus salmonis* Krøyer): migration, growth, and mortality. *Canadian Journal of Fisheries and Aquatic Sciences*, **53**(12), 2808–2813.

Blakey, C.A., & Litt, M.D. (2016). Epigenetic gene expression – An introduction. In: S. Huang, M.D. Litt & C.A. Blakey. *Epigenetic gene regulation and regulation.* London, San Diego: Elsevier. pp. 1-16.

Blanquart, F., Gandon, S. & Nuismer, S. L., (2012). The effects of migration and drift on local adaptation to a heterogeneous environment. *Journal of Evolutionary Biology*, **25**: 1351–1363.

Bonham, K., Zafarullah, M., & Gedamu, L. (1987). The rainbow trout metallothioneins: molecular cloning and characterization of two distinct cDNA Sequences. *DNA*, **6**(6), 519–528.

Bougas, B., Normandeau, E., Pierron, F., Campbell, P. G. C., Bernatchez, L., & Couture, P. (2013). How does exposure to nickel and cadmium affect the transcriptome of yellow perch (*Perca flavescens*) – Results from a 1000 candidate-gene microarray. *Aquatic Toxicology*, **142**: 355–364.

Bourret, V., Couture, P., Campbell, P. G. C., & Bernatchez, L. (2008). Evolutionary ecotoxicology of wild yellow perch (*Perca flavescens*) populations chronically exposed to a polymetallic gradient. *Aquatic Toxicology*, **86**(1), 76–90.

Bouza, C., Vilas, R., Castro, J., & Martínez, P. (2008). Mitochondrial haplotype variability of brown trout populations from Northwestern Iberian Peninsula, a secondary contact area between lineages. *Conservation Genetics*, **9**(4): 917–920.

Bowell, R. J., & Bruce, I. (1995). Geochemistry of iron ochres and mine waters from Levant Mine, Cornwall. *Applied Geochemistry*, **10**(2), 237–250.

Braña, F., Nicieza, A. G., & Toledo, M. M. (1992). Effects of angling on population structure of brown trout, *Salmo trutta* L., in mountain streams of Northern Spain. *Hydrobiologia*, **237**(1), 61–66.

Bridcut, E. E., & Giller, P. S. (1995). Diet variability and foraging strategies in brown trout (*Salmo trutta*): an analysis from subpopulations to individuals. *Canadian Journal of Fisheries and Aquatic Sciences*, **52**(12), 2543–2552.

Bridges, C. C., & Zalups, R. K. (2005). Molecular and ionic mimicry and the transport of toxic metals. *Toxicology and Applied Pharmacology*, **204**(3), 274–308.

Brown, A. R., Gunnarsson, L., Kristiansson, E., & Tyler, C. R. (2014). Assessing variation in the potential susceptibility of fish to pharmaceuticals, considering evolutionary differences in their physiology and ecology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **369**(1656), 20130576–20130576.

Brown, B. E. (1977). Effects of mine drainage on the River Hayle, Cornwall a) factors affecting concentrations of copper, zinc and iron in water, sediments and dominant invertebrate fauna. *Hydrobiologia*, **52**(2-3), 221–233.

Brown, M.E. (1946). The growth of brown trout (*Salmo trutta* L.). I. Factors influencing the growth of brown trout fry. *Journal of Experimental Biology*, **22**: 118-29.

Bryan, G. W. & Hummerstone, L. G. (1977). Indicators of heavy-metal contamination in the Looe Estuary (Cornwall) with particular regard to silver and lead. *Journal of the Marine Biological Association of the United Kingdom*, **57**(1), 75.

Buckley, J.A. (2012). *The Cornish mining industry: a brief history.* Cornwall: Tor Mark Press.

Burt, R., Waite, P. & Burnle Y, R. (1987). *Cornish Mines*. Exeter: University of Exeter Press, pp. 562

Butler, J. R. A., Radford, A., Riddington, G., & Laughton, R. (2009). Evaluating an ecosystem service provided by Atlantic salmon, sea trout and other fish species in the River Spey, Scotland: The economic impact of recreational rod fisheries. *Fisheries Research*, **96**(2), 259–266.

Butler, J.R.A. & Watt, J. (2003) Assessing and managing the impacts of marine salmon farms on wild Atlantic salmon in western Scotland: Identifying priority rivers for conservation. In: D. Mills (ed) *Salmon at the Edge.* Oxford: Blackwell Science Ltd. pp. 93-118.

Byrne, C. J., Poole, R., Dillane, M., Rogan, G., & Whelan, K. F. (2004). Temporal and environmental influences on the variation in sea trout (*Salmo trutta* L.) smolt migration in the Burrishoole system in the west of Ireland from 1971 to 2000. *Fisheries Research*, **66**(1), 85–94.

Camm, G. S., Glass, H. J., Bryce, D. W., & Butcher, A. R. (2004). Characterisation of a mining-related arsenic-contaminated site, Cornwall, UK. *Journal of Geochemical Exploration*, **82**(1), 1–15.

Camm, S. (2010). *Cornish rocks and minerals.* Cornwall: Alison Hodge.

Campbell, D. & Bernatchez, 2004. Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Molecular Biology Evolution,* **21**(5): 945-956.

Campbell, R. N. (1979), Ferox trout, *Salmo trutta* L., and char, *Salvelinm alpinus* (L.), in Scottish lochs. *Journal of Fish Biology*, **14**: 1–29.

Campos, J. L., Posada, D., & Morán, P. (2006). Genetic variation at MHC, mitochondrial and microsatellite loci in isolated populations of Brown trout (*Salmo trutta*). *Conservation Genetics*, **7**(4), 515–530.

Cattanéo, F., Lamouroux, N., Breil, P., & Capra, H. (2002). The influence of hydrological and biotic processes on brown trout (*Salmo trutta*) population dynamics. *Canadian Journal of Fisheries and Aquatic Sciences*, **59**(1), 12–22.

Chiverrell, R. C., & Thomas, G. S. P. (2010). Extent and timing of the Last Glacial Maximum (LGM) in Britain and Ireland: A review. *Journal of Quaternary Science*, **25**(4), 535–549.

Chowdhury, M. J., Baldisserotto, B., & Wood, C. M. (2005). Tissue-specific cadmium and metallothionein levels in rainbow trout chronically acclimated to waterborne or dietary cadmium. *Archives of Environmental Contamination and Toxicology*, **48**(3), 381–390.

Clark, P.U., Dyke, A.S., Shakun, J.D., Carlson, A.E., Clark, J., Wohlfarth, B., Mitrovica, J.X., Hostetler, S.W., & McCabe, A.M. (2009). The Last Glacial Maximum. *Science*, **325**(5941): 710–714.

Coard, R., & Chamberlain, A. T. (1999). The nature and timing of faunal change in the British Isles across the Pleistocene/Holocene transition. *The Holocene*, **9**(3), 372–376.

Collard, B. C., & Mackill, D. J. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **363**(1491), 557-72.

Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., & Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica*, **142**(1–2), 169–196.

Collingwood, R.G., & Myers, J.N.L., (1937). *Roman Britain and English settlements.* Cheshire, UK: Biblo & Tannen Publishers

Consuegra, S., García De Leániz, C., Serdio, A., González Morales, M., Straus, L. G., Knox, D. and Verspoor, E. (2002), Mitochondrial DNA variation in Pleistocene and modern Atlantic salmon from the Iberian glacial refugium. *Molecular Ecology*, **11**: 2037–2048.

Costello, M. J. (2009). How sea lice from salmon farms may cause wild salmonid declines in Europe and North America and be a threat to fishes elsewhere.

Proceedings of the Royal Society of London B: Biological Sciences, **276**(1672), 3385-3394.

Coulthard, T. J., & Macklin, M. G. (2003). Modeling long-term contamination in river systems from historical metal mining. *Geology*, **31**(5), 451.

Crandall, K.A., Bininda-Emonds, O.R., Mace, G.M. & Wayne, R.K. (2000). Considering evolutionary processes in conservation biology. *Trends in Ecology & Evolution*, **15**(7), 290–295.

Crespi, B. J., & Fulton, M. J. (2004). Molecular systematics of Salmonidae: combined nuclear data yields a robust phylogeny. *Molecular Phylogenetics and Evolution*, **31**(2), 658–679.

Crête-Lafrenière, A., Weir, L. K., Bernatchez, L. (2012). Framing the Salmonidae family phylogenetic portrait: A more complete picture from increased taxon sampling. *PLoS ONE*, **7**(10), e46662.

Crisp, D.T. (2000). *Trout and salmon: Ecology, conservation and rehabilitation.* Oxford: Fishing New Books.

Crivelli, A.J. 2006. *Salmo marmoratus*. The IUCN Red List of Threatened Species 2006: e.T19859A9043279. Downloaded on 14 Nov 2016.

Crowl, T. A., Townsend, C. R., & Mcintosh, A. R. (1992). The impact of introduced brown and rainbow trout on native fish: the case of Australasia. *Reviews in Fish Biology and Fisheries*, **2**(3), 217–241.

Dayan, D. I., Crawford, D. L., & Oleksiak, M. F. (2015). Phenotypic plasticity in gene expression contributes to divergence of locally adapted populations of *Fundulus heteroclitus*. *Molecular Ecology*, **24**(13), 3345–3359. De Wit, P. & Palumbi, S. R., 2013. Transcriptome-wide polymorphisms of red abalone (*Haliotis rufescens*) reveal patterns of gene flow and local adaptation. *Molecular Ecology*, **22**(11): 2884–2897.Defo, M. A., Bernatchez, L., Campbell, P. G. C., & Couture, P. (2015). Transcriptional and biochemical markers in transplanted *Perca flavescens* to characterize cadmium- and copper-induced oxidative stress in the field. *Aquatic Toxicology*, **162**, 39–53.

Dekkers, J. C. M. (2004). Commercial application of marker- and gene-assisted selection in livestock: Strategies and lessons. *Journal of Animal Science*, **82**(E_suppl), E313–E328.

Dekkers, J.C.M. & Hospital, F., (2002). Multifactorial genetics: The use of molecular genetics in the improvement of agricultural populations. *Nature Reviews Genetics*, **3**: 22-32.

Denic, M., & Geist, J. (2009). Habitat suitability analysis for lacustrine brown trout (*Salmo trutta*) in Lake Walchensee, Germany: implications for the conservation of an endangered flagship species. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **20**(1), 9–17.

DeWitt, T.J. & Scheiner, S.M. (2004). Phenotypic plasticity: functional and conceptual Approaches. Oxford University Press, Oxford.

DeWitt, T. J., Sih, A., & Wilson, D. S. (1998). Costs and limits of phenotypic plasticity. *Trends in Ecology & Evolution*, **13**(2), 77–81.

Dhuyvetter, H., Gaublomme, E. & Desender, K., 2004. Genetic differentiation and local adaptation in the salt-marsh beetle *Pogonus chalceus*: a comparison between allozyme and microsatellite loci. *Molecular Ecology*, **13**(5): 1065–1074.

Dines, H.G. (1956). The metalliferous mining region of South-west England. Volume II. HMSO. London.

Dionne, M., Caron, F., Dodson, J. J., & Bernatchez, L. (2008). Landscape genetics and hierarchical genetic structure in Atlantic salmon: the interaction of gene flow and local adaptation. *Molecular Ecology*, **17**(10), 2382–2396.

Dorn, L. A., Hammond Pyle, E., & Schmitt, J. (2000). Plasticity to light cues and resources in *Arabidopsis thaliana*: testing for adaptive value and costs. *Evolution*, **54**(6), 1982-94.

Driessnack, M. K., Matthews, A. L., Raine, J. C., & Niyogi, S. (2016). Interactive effects of chronic waterborne copper and cadmium exposure on tissue-specific metal accumulation and reproduction in fathead minnow (*Pimephales promelas*). *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, **179**, 165–173.

Dubois, P. C. A., Trynka, G., Franke, L., Hunt, K. A., Romanos, J., Curtotti, A., *et al.* (2010). Multiple common variants for celiac disease influencing immune gene expression. *Nature Genetics*, **42**(4), 295–302.

Durrant, C. J., Stevens, J. R., Hogstrand, C., & Bury, N. R. (2011). The effect of metal pollution on the population genetic structure of brown trout (*Salmo trutta* L.) residing in the River Hayle, Cornwall, UK. *Environmental Pollution,* **159**(12): 3595–603.

Edge, S. E., Morgan, M. B., Gleason, D. F., & Snell, T. W. (2005). Development of a coral cDNA array to examine gene expression profiles in *Montastraea faveolata* exposed to environmental stress. *Marine Pollution Bulletin*, **51**(5), 507–523.

Edwards, R. W., Densem, J. W., & Russell, P. A. (1979). An assessment of the importance of temperature as a factor controlling the growth rate of brown trout in streams. *The Journal of Animal Ecology*, **48**(2), 501.

Ehlen, J. (2005). Effects of human activity on lineation analysis. *Reviews in Engineering Geology XVI: Humans as Geologic Agents* **16**, 7–18.

Ehlen, J., Haneberg, W.C., & Larson, R.A. (eds.) (2005). *Humans as geologic agents*. Boulder, Colorado: The Geological Society of America.

Eizaguirre, C., Lenz, T. L., Kalbe, M., Milinski, M. (2012), Divergent selection on locally adapted major histocompatibility complex immune genes experimentally proven in the field. *Ecology Letters*, **15**(7): 723–731.

Eizaguirre, C., Lenz, T.L., Sommerfeld, R.D., Harrod, C., Kalbe, M. & Milinski, M., 2011. Parasite diversity, patterns of MHC II variation and olfactory based mate choice in diverging three-spined stickleback ecotypes. *Evolutionary Ecology*, **25**(3): 605-622.

Elliott, J. M. (1976). The energetics of feeding, metabolism and growth of brown trout (*Salmo trutta* L.) in relation to body weight, water temperature and ration size. *The Journal of Animal Ecology*, **45**(3), 923.

Elliott, J. M., & Elliott, J. A. (2010). Temperature requirements of Atlantic salmon *Salmo salar*, brown trout *Salmo trutta* and Arctic charr *Salvelinus alpinus*: predicting the effects of climate change. *Journal of Fish Biology*, **77**(8), 1793–1817.

Elliott, J.M. (1994). *Quantitative ecology and the brown trout*. Oxford Series in Ecology and Evolution, Oxford: Oxford University Press.

Elo, K., Ivanoff, S., Vuorinen, J. A., & Piironen, J. (1997). Inheritance of RAPD markers and detection of interspecific hybridization with brown trout and Atlantic salmon. *Aquaculture*, **152**(1-4), 55–65.

Environment Agency (2009). Economic evaluation of inland fisheries. Mangers report from science project SC050026/SR2.

Environment Agency (2013). Abandoned mines and the water environment. Science Project: SC030136-41.

Environment Agency. (2008a). Evaluating options for sea trout and brown trout biological reference points. Science Report SC060070.

Environment Agency. (2008b).The thermal biology of brown trout and Atlantic salmon. Science Summary: SCHO1008BOUE-E-P.

Epstein, B., Jones, M., Hamede, R., Hendricks, S., McCallum, H., Murchison, E. P., Schönfeld, B., Wiench, C., Hohenlohe, P. & Storfer, A. (2016). Rapid evolutionary response to a transmissible cancer in Tasmanian devils. *Nature Communications*, **7**(12684), 1-7.

Ergon, T., Lambin, X., & Stenseth, N. C. (2001). Life-history traits of voles in a fluctuating population respond to the immediate environment. *Nature*, **411**(6841), 1043–1045.

Etter, P. D., Preston, J. L., Bassham, S., Cresko, W. A, & Johnson, E. A. (2011). Local de novo assembly of RAD paired-end contigs using short sequencing reads. *PloS One*, **6**(4), e18561.

Evans, M.L., Neff, B.D. & Heath, D.D., 2010. MHC-mediated local adaptation in reciprocally translocated Chinook salmon. *Conservation Genetics*, **11**(6): 2333-2342.

Fatima, M., Usmani, N., & Hossain, M. (2014). Heavy metal in aquatic ecosystem emphasizing its effect on tissue bioaccumulation and histopathology: A review. *Journal of Environmental Science and Technology,* **7**(1)*: 1-15.*

Fausch, K. D., & White, R. J. (1981). Competition between brook trout (*Salvelinus fontinalis*) and brown trout (*Salmo trutta*) for positions in a Michigan stream. *Canadian Journal of Fisheries and Aquatic Sciences*, **38**(10), 1220–1227.

Feil, R., & Fraga, M. F. (2012). Epigenetics and the environment: emerging patterns and implications. *Nature Reviews Genetics*, **13**(2), 97-109.

Ferguson, A. (1989). Genetic differences among brown trout, *Salmo trutta*, stocks and their importance for the conservation and management of the species. *Freshwater Biology*, **21**(1), 35–46.

Ferguson, A., (2006). Genetics of sea trout, with particular reference to Britain and Ireland. In: Harris, G. & Milner, N. (eds.) *Sea trout: biology, conservation and management: Proceedings of the First International Sea Trout Symposium, Cardiff, July 2004.* Oxford: Blackwell Publishing Ltd., pp. 157-182.

Ferguson, A., & Mason, F. M. (1981). Allozyme evidence for reproductively isolated sympatric populations of brown trout *Salmo trutta* L. in Lough Melvin, Ireland. *Journal of Fish Biology*, **18**(6), 629–642.

Ferguson, A., & Taggart, J. B. (1991). Genetic differentiation among the sympatric brown trout (*Salmo trutta*) populations of Lough Melvin, Ireland. *Biological Journal of the Linnean Society*, **43**(3), 221–237.

Finch, C. E. (1998). Variations in senescence and longevity include the possibility of negligible senescence. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, **53**(4), B235-9.

Finnegan, A. K., Griffiths, A. M., King, R. A., Machado-Schiaffino, G., Porcher, J.-P., Garcia-Vazquez, E., Bright, D., & Stevens, J. R. (2013). Use of multiple markers demonstrates a cryptic western refugium and postglacial colonisation routes of Atlantic salmon (*Salmo salar* L.) in northwest Europe. *Heredity*, **111**(1), 34–43.

Fischer, M.C., Foll, M., Excoffier, L., & Heckel, G., 2011. Enhanced AFLP genome scans detect local adaptation in high-altitude populations of a small rodent (*Microtus arvalis*). *Molecular Ecology*, **20**(7): 1450–1462.

Flatscher, R., Frajman, B., Schönswetter, P., & Paun, O. (2012). Environmental heterogeneity and phenotypic divergence: can heritable epigenetic variation aid speciation? *Genetics Research International*, **2012**, 1–9.

Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**(2), 977–993.

Foust, C. M., Preite, V., Schrey, A. W., Alvarez, M., Robertson, M. H., Verhoeven, K. J. F., & Richards, C. L. (2016). Genetic and epigenetic differences associated with environmental gradients in replicate populations of two salt marsh perennials. *Molecular Ecology*, **25**(8), 1639–1652.

Frank-Gopolos, T., Bekkevold, D., Guyomard, R., De Innocentiis, S., Martínez Portela, P., Fernández, C.B., & Vera, M. (2015). Aquatrace report for brown trout. Aquatrace.

Fraser, D. J. and Bernatchez, L. (2001), Adaptive evolutionary conservation: towards a unified concept for defining conservation units. *Molecular Ecology*, **10**: 2741–2752.

Fraser, D. J., Weir, L. K., Bernatchez, L., Hansen, M. M., & Taylor, E. B. (2011). Extent and scale of local adaptation in salmonid fishes: review and meta-analysis. *Heredity*, **106**(3), 404–420.

Freeman, S. & Heron, S. (2004). *Evolutionary Analysis.* Third Edition. Cambridge: Pearson.

Freyhof, J., (2011). *Salmo trutta*. The IUCN Red List of Threatened Species 2011: e.T19861A9050312. Downloaded on 14 November 2016.

Galvez, F., & Wood, C. (2002). The mechanisms and costs of physiological and toxicological acclimation to waterborne silver in juvenile rainbow trout (*Oncorhynchus mykiss*). *Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology*, **172**(7), 587–597.

Gamble, B. (2011). Cornish Mines: St Just to Redruth. Cornwall: Alison Hodge.

Garcia de Leaniz, C., Fleming, I.A., Einum, S., Verspoor, E., Jordan, W.C., Consuegra, S. Aubin-Horth, N., Lajus, D., Letcher, B.H., Youngson, A.F., Webb, J.H., Vøllestad, L.A., Villanueva, B., Ferguson, A., & Quinn, T.P. (2007). A critical review of adaptive

genetic variation in Atlantic salmon: implications for conservation. *Biol. Rev.,* **82**: 173–211.

GarcÍa-MarÍn, J.-L., Utter, F. M., & Pla, C. (1999). Postglacial colonization of brown trout in Europe based on distribution of allozyme variants. *Heredity*, **82**(1), 46–56.

Gene Ontology Consortium. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research*, **32**(90001), 258D–261.

Gerrard, S. (2000). The Early British tin Industry. *Tempus Publishing Ltd,* Stroud.

Gharbi, K., Gautier, A., Danzmann, R. G., Gharbi, S., Sakamoto, T., Høyheim, B., Taggart, J.B., Cairney, M., Powell, R., Krieg, F., Okamoto, N., Ferguson, M.M., Holm, L-E., & Guyomard, R. (2006). A linkage map for brown trout (*Salmo trutta*): chromosome homeologies and comparative genome organization with other salmonid fish. *Genetics*, **172**(4): 2405–19.

Glasauer, S. M. K., & Neuhauss, S. C. F. (2014). Whole-genome duplication in teleost fishes and its evolutionary consequences. *Molecular Genetics and Genomics*, **289**(6), 1045–1060.

Goddard, M. E., & Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics*, **10**(6), 381–391.

Goodwin, J. C. A., R. A. King, Jones, I.J., Ibbotson, A., & Stevens, J. R. (2016). A small number of anadromous females drive reproduction in a brown trout (*Salmo trutta*) population in an English chalk stream. *Freshwater Biology*, **61**(7), 1075–1089.

Goody, J. (2012). Metals, culture and capitalism: An essay on the origins of the modern world. Cambridge: Cambridge University Press.

Gosset, C., Rives, J., & Labonne, J. (2006). Effect of habitat fragmentation on spawning migration of brown trout (*Salmo trutta* L.). *Ecology of Freshwater Fish*, **15**(3), 247–254.

Greig, S. M., Sear, D. A. and Carling, P. A. (2007). A review of factors influencing the availability of dissolved oxygen to incubating salmonid embryos. *Hydrol. Process.*, **21**: 323–334.

Griffiths, A. (2005) *The population structure of Brown Trout (Salmo trutta) on Dartmoor National Park.* PhD Thesis. University of Exeter

Griffiths, A. M., Ellis, J. S., Clifton-Dey, D., Machado-Schiaffino, G., Bright, D., Garcia-Vazquez, E., & Stevens, J. R. (2011). Restoration versus recolonisation: The origin of Atlantic salmon (*Salmo salar* L.) currently in the River Thames. *Biological Conservation*, **144**(11), 2733-2738.

Griffiths, A. M., Koizumi, I., Bright, D., & Stevens, J. R. (2009). A case of isolation by distance and short-term temporal stability of population structure in brown trout (*Salmo trutta*) within the River Dart, southwest England. *Evolutionary Applications*, **2**(4), 537–554.

Grosell, M., Nielsen, C., & Bianchini, A. (2002). Sodium turnover rate determines sensitivity to acute copper and silver exposure in freshwater animals. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, **133**(1), 287–303.

Gunnarsson, L., Kristiansson, E., Förlin, L., Nerman, O., & Joakim Larsson, D.G. (2007). Sensitive and robust gene expression changes in fish exposed to estrogen – a microarray approach. *BMC Genomics*, **8**(1), 149. **DOI:** 10.1186/1471-2164-8-149

Gyllensten, U. (1985). The genetic structure of fish: differences in the intraspecific distribution of biochemical genetic variation between marine, anadromous, and freshwater species. *Journal of Fish Biology*, **26**(6), 691–699.

Hamilton, K. E., Ferguson, A., Taggart, J. B., Tómasson, T., Walker, A. and Fahy, E. (1989), Post-glacial colonization of brown trout, *Salmo trutta* L.: *Ldh-5* as a phylogeographic marker locus. *Journal of Fish Biology*, **35**: 651–664.

Hamilton, P. B., Cowx, I. G., Oleksiak, M. F., Griffiths, A. M., Grahn, M., Stevens, J. R., Carvalho, G.R., Nicol, E., & Tyler, C. R. (2016). Population-level consequences for wild fish exposed to sublethal concentrations of chemicals - a critical review. *Fish and Fisheries*, **17**(3), 545–566.

Hamilton, S. J., & Mehrle, P. M. (1986). Metallothionein in fish: Review of Its importance in assessing stress from metal contaminants. *Transactions of the American Fisheries Society*, **115**(4), 596–609.

Hansen, M. M., Ruzzante, D. E., Nielsen, E. E., & Mensberg, K.-L. D. (2001). Brown trout (*Salmo trutta*) stocking impact assessment using microsatellite dna markers. *Ecological Applications*, **11**(1), 148–160.

Hansen, M. M., Ruzzante, D. E., Nielsen, E. E., Bekkevold, D. and Mensberg, K.-L. D. (2002). Long-term effective population sizes, temporal stability of genetic composition and potential for local adaptation in anadromous brown trout (*Salmo trutta*) populations. *Molecular Ecology*, **11**: 2523–2535.

Hao, W., & Golding, G. B. (2006). The fate of laterally transferred genes: life in the fast lane to adaptation or death. *Genome Research*, **16**(5), 636–43.

Harris, T.R. (1945) Engineering in Cornwall before 1775, Transactions of the Newcomen Society, **25**(1): 111-122.

Hartl, D.L. & Jones, E.W. (2009) *Essential Genetics: A genomics perspective*. Fifth Edition. Boston, USA: Jones & Bartlett Learning

Hatcher, J. (1973) English tin Production and Trade Before 1550. Oxford: Clarendon Press.

Hayama, R., Yokoi, S., Tamaki, S., Yano, M., & Shimamoto, K. (2003). Adaptation of photoperiodic control pathways produces short-day flowering in rice. *Nature*, **422**(6933), 719–722.

Heath, A.G. (1995). *Water pollution and fish physiology*. Second edition. Boca Raton, New York, London, Tokyo: Lewis Publishers

Hecht, B. C., Campbell, N. R., Holecek, D. E., & Narum, S. R. (2013). Genome-wide association reveals genetic basis for the propensity to migrate in wild populations of rainbow and steelhead trout. *Molecular Ecology*, **22**(11), 3061–3076.

Herbert, D. W., & Merkens, J. C. (1961). The effect of suspended mineral solids on the survival of trout. *International Journal of Air and Water Pollution*, **5**, 46–55.

Hoffmann, A. A., & Hercus, M. J. 2000. Environmental stress as an evolutionary force. *BioScience,* **50**(3)*217–226.*

Hoffmann, A.A. & Willi, Y., 2008. Detecting genetic responses to environmental change. *Nature Reviews Genetics*, 9**:** 217-226.

Hogan, D.L., & Church, M. (1989). Hydraulic geometry in small, coastal streams: progress towards quantification of salmonid habitat. *Canadian Journal of Fisheries and Aquatic Sciences*, **46**: 844-852.

Hogan, D.L., Bird, S.A. & Hassan, M.A. (1998). Spatial and temporal evolution of small coastal gravel-bed streams: influence of forest management on channel morphology and fish habitats. In: P.C. Klingeman, R.L. Beschta, P.C. Komer and J.B. Bradley. *Gravel-bed rivers in the environment*. Colorado, USA: Water Resources Publications, LLC.

Hohenlohe, P. A., Bassham, S., Currey, M., & Cresko, W. A. (2011). Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **367**(1587).

Homer, R.F. (1991). Tin, Lead and Pewter. In J Blair & N Ramsay *English Medieval Industries*. London, New York: The Hambledon Press. pp. 57-81.

Huang, Q. Q., Pan, X. Y., Fan, Z. W., & Peng, S. L. (2015). Stress relief may promote the evolution of greater phenotypic plasticity in exotic invasive species: a hypothesis. *Ecology and Evolution*, **5**(6), 1169–1177.

Hubby, J.L., & Lewontin, R.C., 1966. A molecular approach to the study of genic heterozygosity in Natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura*. Genetics, **54**(2): 577-594.

Hyvarinen, P., & Huusko, A. (2006). Diet of brown trout in relation to variation in abundance and size of pelagic fish prey. *Journal of Fish Biology*, **68**(1), 87–98.

Ikediashi, C., Billington, S., & Stevens, J. R. (2012). The origins of Atlantic salmon (*Salmo salar* L.) recolonizing the River Mersey in northwest England. *Ecology and Evolution*, **2**(10), 2537–48.

Jablonka, E., & Lamb, M. J. (2002). The changing concept of epigenetics. *Annals of the New York Academy of Sciences*, **981**, 82–96.

Jacobi, R. M., Rose, J., MacLeod, A., & Higham, T. F. G. (2009). Revised radiocarbon ages on woolly rhinoceros (*Coelodonta antiquitatis*) from western central Scotland:

significance for timing the extinction of woolly rhinoceros in Britain and the onset of the LGM in central Scotland. *Quaternary Science Reviews*, **28**(25), 2551–2556.

Jaramillo-Correa, J. P., Beaulieu, J. & Bousquet, J., 2001. Contrasting evolutionary forces driving population structure at expressed sequence tag polymorphisms, allozymes and quantitative traits in white spruce. *Molecular Ecology*. **10**(11): 2729–2740.

Jaramillo-Correa, J.-P., Rodríguez-Quilón, I., Grivet, D., Lepoittevin, C., Sebastiani, F., Heuertz, M., Garnier-Géré P.H., Alía, R., Polmion, C., Vendramin G.G., González-Martínez, S. C. (2015). Molecular proxies for climate maladaptation in a long-lived tree (*Pinus pinaster* Aiton, Pinaceae). *Genetics*, **199**(3), 793-807.

Jensen, H., Kiljunen, M., & Amundsen, P.-A. (2012). Dietary ontogeny and niche shift to piscivory in lacustrine brown trout *Salmo trutta* revealed by stomach content and stable isotope analyses. *Journal of Fish Biology*, **80**(7), 2448–2462.

Jensen, L. F., M. M. Hansen, C. Pertoldi, G. Holdensgaard, K. L. Mensberg, & Loeschcke, V. (2008). Local adaptation in brown trout early life-history traits: implications for climate change adaptability. *Proceedings of the Royal Society B: Biological Sciences,* **275**, 2859–2868.

Johansson, N., Larsson, Å., & Lewander, K. (1972). Metabolic effects of PCB (polychlorinated biphenyls) on the brown trout (*Salmo trutta*). *Comparative and General Pharmacology*, **3**(11), 310–314.

Johnson, K. R., Wright, J. E., May, B., & May, B. (1987). Linkage relationships reflecting ancestral tetraploidy in salmonid fish. *Genetics*, **116**(4): 579–91.

Johnston, S. E., Orell, P., Pritchard, V. L., Kent, M. P., Lien, S., Niemelä, E., Erkinaro, J., & Primmer, C. R. (2014). Genome-wide SNP analysis reveals a genetic basis for sea-age variation in a wild population of Atlantic salmon (*Salmo salar*). *Molecular Ecology*, **23**(14), 3452–3468.

Jonsson, B. (1985). Life history patterns of freshwater resident and sea-run migrant brown trout in Norway. *Transactions of the American Fisheries Society*, **114**(2), 182–194.

Jonsson, B., & Jonsson, N. (2009). A review of the likely effects of climate change on anadromous Atlantic salmon *Salmo salar* and brown trout *Salmo trutta*, with particular reference to water temperature and flow. *Journal of Fish Biology*, **75**(10), 2381–2447.

Jonsson, B., & Jonsson, N. (2011). *Ecology of Atlantic salmon and Brown trout: Habitat as a template for life histories*. Dordrecht, Heidelberg, London, New York: Springer.

Jonsson, N., Naesje, T. F., Jonsson, B., Saksgard, R., & Sandlund, O. T. (1999). The influence of piscivory on life history traits of brown trout. *Journal of Fish Biology*, **55**(6), 1129–1141.

Jozefczak, M., Remans, T., Vangronsveld, J., & Cuypers, A. (2012). Glutathione is a key player in metal-induced oxidative stress defenses. *International Journal of Molecular Sciences*, **13**(12), 3145–3175.

Jump, A. S., & Peñuelas, J., 2005. Running to stand still: adaptation and the response of plants to rapid climate change. *Ecology Letters*, **8**(9): 1010–1020.

Kassahn, K. S., Caley, M. J., Ward, A. C., Connolly, A. R., Stone, G., & Crozier, R. H. (2007). Heterologous microarray experiments used to identify the early gene response to heat stress in a coral reef fish. *Molecular Ecology*, **16**(8), 1749–1763.

Kawecki, T.J. & Ebert, D., 2004. Conceptual issues in local adaptation. *Ecology Letters,* **7** (12): 1225-1241.

Kelly, A., Charman, D. J., & Newnham, R. M. (2010). A Last Glacial Maximum pollen record from Bodmin Moor showing a possible cryptic northern refugium in southwest England. *Journal of Quaternary Science*, **25**(3), 296–308.

Kemp, P., Sear, D., Collins, A., Naden, P., & Jones, I. (2011). The impacts of fine sediment on riverine fish. *Hydrological Processes*, **25**(11), 1800–1821.

King, R. A., Hillman, R., Elsmere, P., Stockley, B., & Stevens, J. R. (2016). Investigating patterns of straying and mixed stock exploitation of sea trout, *Salmo trutta,* in rivers sharing an estuary in south-west England. *Fisheries Management and Ecology*, **23**(5), 376–389.

Klerks, P.L. & Levinton, J.S., (1989). Effects of heavy metals in a polluted aquatic ecosystem. pp. 41-67, In: Ecotoxicology: Problems and Approaches, S.A. Levin, M.A. Harwell, J.R. Kelly and K.D. Kimball (Eds.), Springer-Verlag, New York.

Klerks, P.L., Xie, L. & Levinton, J.S., (2011). Quantitative genetics approaches to study evolutionary processes in ecotoxicology; a perspective from research on the evolution of resistance. *Ecotoxicology*, **20**(3): 512-523.

Kondolf, G.M., & Piégay, H. (2003). Tools in fluvial geomorphology. Oxford: Wiley-Blackwell.

Kramer, H. J., Gonick, H. C., & Lu, E. (1986). In vitro inhibition of Na^+K^+-ATPase by trace metals: relation to renal and cardiovascular damage. *Nephron*, **44**(4), 329–36.

Krasnov, A., Koskinen, H., Pehkonen, P., Rexroad, C. E., Afanasyev, S. and Mölsä, H., (2005). Gene expression in the brain and kidney of rainbow trout in response to handling stress. *BMC Genomics*, **6**(1), 3.

Kulac, B., Atli, G., & Canli, M. (2013). Response of ATPases in the osmoregulatory tissues of freshwater fish *Oreochromis niloticus* exposed to copper in increased salinity. *Fish Physiology and Biochemistry*, **39**(2), 391–401.

L'Abee-Lund, J. H., Jonsson, B., Jensen, A. J., Saettem, L. M., Heggberget, T. G., Johnsen, B. O., & Naesje, T. F. (1989). Latitudinal Variation in Life-History Characteristics of Sea-Run Migrant Brown Trout *Salmo trutta*. *The Journal of Animal Ecology*, **58**(2), 525.

L'Abee-Lund, J. H., Langeland, A., & Saegrov, H. (1992). Piscivory by brown trout Salmo trutta L. and Arctic charr Salvelinus alpinus (L.) in Norwegian lakes. *Journal of Fish Biology*, **41**(1), 91–101.

Laikre, L.*,* Antunes, A.*,* Apostolidis, A.*,* Berrebi, P.*,* Duguid, A.*,* Ferguson, A.*,* Garcia-Marin, J.L.*,* Guyomard, R.*,* Hansen, M.M.*,* Hindar, K.*,* Koljonen, M.-L.*,* Lardiadèr, C.*,* Martinez, P.*,* Nielsen, E.E.*,* Palm, S.*,* Ruzzante, D.*,* Ryman, N. *&* Triantaphyllidis, C. *(*1999*)* Conservation genetic management of brown trout (*Salmo trutta*) in Europe*. Report by the concerted action on identification, management and exploitation of genetic resources in the brown trout (Troutconcert EU FAIR CT97-3882).*

Lancaster, L. T., Dudaniec, R. Y., Chauhan, P., Wellenreuther, M., Svensson, E. I., & Hansson, B. (2016). Gene expression under thermal stress varies across a geographical range expansion front. *Molecular Ecology*, **25**(5), 1141–1156.

Lande, R., & Thompson, R. (1990). Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics*, **124**(3), 743-756.

Larsson, Å., Haux, C., & Sjöbeck, M.-L. (1985). Fish physiology and metal pollution: Results and experiences from laboratory and field studies. *Ecotoxicology and Environmental Safety*, **9**(3), 250–281.

Le Corre, V. & Kremer, A., 2012. The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology*, **21**(7): 1548–1566.

Leadbeater, C. (2012). We think – mass innovation not mass production. Exmouth, UK: Profile Books Ltd.

LeBoutillier, N.G. (2002). The tectonics of Variscan magmatism and mineralisation in Southwest England. Volume II. PhD thesis. University of Exeter

Lee JR, Rose J, Hamblin RJO, Moorlock, B.S.P., Riding, J.B., Phillips, E., Barendregt, R.W., & Candy, I. (2011). The glacial history of the British Isles during the Early and Middle Pleistocene: implications for the long-term development of the British Ice Sheet. Developments in Quaternary Sciences **15***:* 59–74

Lee, J.-Y., & Lee, D.-H. (2003). Use of serial analysis of gene expression technology to reveal changes in gene expression in *Arabidopsis* pollen undergoing cold stress. *Plant Physiology*, **132**(2), 517–29.

Leifchild, J.R. (1968). *Cornwall, its mines and miners*. New York: Routledge.

Lerceteau-Köhler *et al.,*(2013). Genetic variation in brown trout *Salmo trutta* across the Danube, Rhine, and Elbe headwaters: a failure of the phylogeographic paradigm? *BMC Evolutionary Biology*, **13**(1), 176. DOI 10.1186/1471-2148-13-176.

Leung, C., Breton, S., & Angers, B. (2016). Facing environmental predictability with different sources of epigenetic variation. *Ecology and Evolution*, **6**(15), 5234–5245.

Lewis, G. R. (1965). *The Stanneries: A Study of Medieval Tin Miners of Cornwall.* D. Bradford Barton Ltd., Truro.

Li, J., Haipeng, L., Jakobsson, M., Sjodin, P. & Lascoux, M., (2012). Joint analysis of demography and selection in population genetics: where do we stand and where could we go? *Molecular Ecology,* **21**(1)*,* 28–44*.*

Lien, S., Gidskehaug *et al.,*(2011). A dense SNP-based linkage map for Atlantic salmon (*Salmo salar*) reveals extended chromosome homeologies and striking differences in sex-specific recombination patterns. *BMC Genomics*, **12**(1), 615.

Lister, A. M., & Stuart, A. J. (2008). The impact of climate change on large mammal distribution and extinction: Evidence from the last glacial/interglacial transition. *Comptes Rendus Geoscience*, **340**(9), 615–620.

Lotterhos, K. E. and Whitlock, M. C. (2014), Evaluation of demographic history and neutral parameterization on the performance of F_{ST} outlier tests. *Molecular Ecology*, **23**: 2178–2192.

Luckenbach, T., Ferling, H., Gernhöfer, M., Köhler, H.-R., Negele, R.-D., Pfefferle, E., & Triebskorn, R. (2003). Developmental and subcellular effects of chronic exposure to sub-lethal concentrations of ammonia, PAH and PCP mixtures in brown trout (*Salmo trutta f. fario L*.) early life stages. *Aquatic Toxicology*, **65**(1), 39–54.

Luikart, G., England, P.R., Tallmon, D., Jordan, S. & Taberlet, P., (2003). The power and promise of population genomics: from genotyping to genome typing. Nature Reviews Genetics, **4**: 981-994.

Lynch, M., & Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sunderland: Sinauer.

MacKenzie, K., Longshaw, M., Begg, G. S., McVicar MacKenzie, A. H., McVicar, A. H., & to MacKenzie, C. K. (1998). Sea lice (Copepoda: Caligidae) on wild sea trout (*Salmo trutta* L.) in Scotland. *ICES Journal of Marine Science*, **55**, 151–162.

Macnair, M. R. (1987). Heavy metal tolerance in plants: A model evolutionary system. *Trends in Ecology & Evolution*, **2**(12), 354–359.

MacPherson, J. M., Eckstein, P. E., Scoles, G. J., & Gajadhar, A. A. (1993). Variability of the random amplified polymorphic DNA assay among thermal cyclers, and effects of primer and DNA concentration. *Molecular and Cellular Probes*, **7**(4), 293–299.

Mangel, M. (1996). Life history invariants, age at maturity and the ferox trout. *Evolutionary Ecology*, **10**(3), 249–263.

Manning, W.H., (1979). The native and Roman contribution to the development of metal industries in Britain. In B.C. Burnham & H.B. Johnson (eds.) *Invasion and response: the case of Roman Britain*, BAR, British Series, LXXIII, Oxford, pp. 111-121.

Manolio, T. A., *et al.,*(2009). Finding the missing heritability of complex diseases. *Nature*, **461**(7265), 747–753.

Martínez, P., Bouza, C., Castro, J., Hermida, M., Pardo, B. G., & Sánchez, L. (2007). Analysis of a secondary contact between divergent lineages of brown trout *Salmo trutta L*. from Duero basin using microsatellites and mtDNA RFLPs. *Journal of Fish Biology*, **71**: 195–213.

Maryoung, L. A., Blunt, B., Tierney, K. B., & Schlenk, D. (2015). Sublethal toxicity of chlorpyrifos to salmonid olfaction after hypersaline acclimation. *Aquatic Toxicology*, **161**, 94–101.

Maynard-Smith, J., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical Research*, **23**(1), 23–35.

McDonald, D. G., & Wood, C. M. (1993). Branchial mechanisms of acclimation to metals in freshwater fish. In *Fish Ecophysiology*. Dordrecht: Springer Netherlands, pp. 297–321.

McHugh, P., Budy, P., Thiede, G., VanDyke, E. (2008). Trophic relationships of nonnative brown trout, *Salmo trutta*, and native Bonneville cutthroat trout, *Oncorhynchus clarkii utah*, in a northern Utah, USA river. *Environ Biol Fish,* **81**: 63-75**.**

McKay, J. K., & Latta, R. G. (2002). Adaptive population divergence: markers, QTL and traits. *Trends in Ecology & Evolution*, **17**(6), 285–291.

McKeown, N. J., Hynes, R. A., Duguid, R. A., Ferguson, A. and Prodöhl, P. A. (2010). Phylogeographic structure of brown trout *Salmo trutta* in Britain and Ireland: glacial refugia, postglacial colonization and origins of sympatric populations. *Journal of Fish Biology*, **76**, 319–347.

McVeigh, H. P., Hynes, R. A., & Ferguson, A. (1995). Mitochondrial DNA differentiation of sympatric populations of brown trout, *Salmo trutta* L., from Lough Melvin, Ireland. *Canadian Journal of Fisheries and Aquatic Sciences*, **52**(8), 1617–1622.

Meador, J. P. (2015). Tissue concentrations as the dose metric to assess potential toxic effects of metals in field-collected fish: Copper and cadmium. *Environmental Toxicology and Chemistry*, **34**(6), 1309–1319.

Meier, K., Hansen, M. M., Bekkevold, D., Skaala, Ø., & Mensberg, K.-L. D. (2011). An assessment of the spatial scale of local adaptation in brown trout (*Salmo trutta L.*): footprints of selection at microsatellite DNA loci. *Heredity*, **106**(3), 488–499.

Meissner, K. & Muotka, T. (2006). The role of trout in stream food webs: integrating evidence from field surveys and experiments. *Journal of Animal Ecology*, **75:** 421–433.

Meland, S., Heier, L. S., Salbu, B., Tollefsen, K. E., Farmen, E., & Rosseland, B. O. (2010). Exposure of brown trout (*Salmo trutta L.*) to tunnel wash water runoff — Chemical characterisation and biological impact. *Science of The Total Environment*, **408**(13), 2646–2656.

Mengoni, A., Gonnelli, C., Galardi, F., Gabbrielli, R. & Bazzicalupo, M., (2000). Genetic diversity and heavy metal tolerance in populations of *Silene paradoxa* L. (Caryophyllaceae): a random amplified polymorphic DNA analysis. *Molecular Ecology,* **9**(9): 1319–1324.

Meraner, A., Baric, S., Pelster, B., & Dalla Via, J. (2010). Microsatellite DNA data point to extensive but incomplete admixture in a marble and brown trout hybridisation zone. *Conservation Genetics*, **11**(3), 985–998.

Merilä, J. and Hendry, A. P. (2014), Climate change, adaptation, and phenotypic plasticity: the problem and the evidence. *Evolutionary Applications*, **7**: 1–14.

Minghetti, M., Schnell, S., Chadwick, M. A., Hogstrand, C., & Bury, N. R. (2014). A primary FIsh Gill Cell System (FIGCS) for environmental monitoring of river waters. *Aquatic Toxicology*, **154**, 184–192.

Mitchell, J. & Moore, S. (2009). Tuckingmill Valley Park Management Plan 2009. Kerrier & Camborne District Councils

Mitton, J. B., & Duran, K. L., (2004). Genetic variation in piñon pine, *Pinus edulis*, associated with summer precipitation. *Molecular Ecology*. **13**(5): 1259–1264.

Moore, L. J., Somamoto, T., Lie, K. K., Dijkstra, J. M., & Hordvik, I. (2005). Characterisation of salmon and trout CD8α and CD8β. *Molecular Immunology*, **42**(10), 1225–1234.

Moritz, C. (1994). Defining "Evolutionarily Significant Units" for conservation. *Trends in Ecology & Evolution*, **9**(10), 373–375.

Murren, C. J., Auld, J. R., Callahan, H., Ghalambor, C. K., Handelsman, C. A., Heskel, M. A., Kingsolver, J.G., Maclean, H.J., Masel, J., Maughan, H., Pfennig, D.W., Relyea, R.A., Seiter, S., Snell-Rood, E., Steiner, U.K., Schlichting, C. D. (2015). Constraints on the evolution of phenotypic plasticity: limits and costs of phenotype and plasticity. *Heredity*, **115**(4), 293–301.

Nadig, S.G., Lee, K.L., Adams, S.M., (1998). Evaluating alterations of genetic diversity in sunfish populations exposed to contaminants using RAPD assay. *Aquatic Toxicology*, **43**(2-3): 163-178.

Narum, S. R., Buerkle, C. A., Davey, J. W., Miller, M. R. & Hohenlohe, P. A., (2013). Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, **22**(11): 2841–2847.

Newman, P. (1998). The Dartmoor tin industry – a field guide. Newton Abbot, Devon, UK. Chercombe Press

Nichols, K. M., Edo, A. F., Wheeler, P. A., & Thorgaard, G. H. (2008). The genetic basis of smoltification-related traits in *Oncorhynchus mykiss*. *Genetics*, **179**(3).

Norris, D. O., Donahue, S., Dores, R. M., Lee, J. K., Maldonado, T. A., Ruth, T., & Woodling, J. D. (1999). Impaired adrenocortical response to stress by brown trout, *Salmo trutta*, living in metal-contaminated waters of the Eagle River, Colorado. *General and Comparative Endocrinology*, **113**(1), 1–8.

Nuvolari, A. (2004). Collective invention during the British Industrial Revolution: the case of the Cornish pumping engine. *Cambridge Journal of Economics*, **28**(3), 347–363.

O'Dea, R. E., Noble, D. W. A., Johnson, S. L., Hesselson, D., & Nakagawa, S. (2016). The role of non-genetic inheritance in evolutionary rescue: epigenetic buffering, heritable bet hedging and epigenetic traps. *Environmental Epigenetics*, **2**(1).

Ohno, S. (1970). Evolution by gene duplication. London: Allen and Unwin.

Oksanen, T. A., Koivula, M., Koskela, E., Mappes, T., & Soulsbury, C. D. (2012). Interactive effects of past and present environments on overwintering success-a reciprocal transplant experiment. *Ecology and Evolution*, **2**(5), 899–907.

Olsson, I. C., Greenberg, L. A., Bergman, E., & Wysujack, K. (2006). Environmentally induced migration: the importance of food. *Ecology Letters*, **9**(6), 645–651.

Olsson, M., & Shine, R. (2002). Growth to death in lizards. *Evolution*, **56**(9), 1867.

Olsson, P-E., Kling, P., Hogstrand, C. (1998). Mechanisms of heavy metal accumulation and toxicity in fish. In W.J. Langston, & M.J. Bebianno. *Metal metabolism in aquatic environments.* London, Weinheim, New York, Tokyo, Melbourne, Madras: Chapman & Hall. pp. 321-350.

Ondarza, P. M., Gonzalez, M., Fillmann, G., & Miglioranza, K. S. B. (2011). Polybrominated diphenyl ethers and organochlorine compound levels in brown trout (*Salmo trutta*) from Andean Patagonia, Argentina. *Chemosphere*, **83**(11), 1597–1602.

Orr, H. A. 2005. The genetic theory of adaptation: a brief history. *Nature Reviews Genetics* **6**:119-127.

Ouborg, N.J., Pertoldi, C., Loeschcke, V., Biklsma, R., Hedrick, P.W., (2010). Conservation genetics in transition to conservation genomics. *Trends in Genetics,* **26**(4): 177-187.

Owen, J. (2012). *Trout.* Reaktion Books Ltd., London. pp. 147.

Parra, I., Almodóvar, A., Ayllón, D., Nicola, G. G., & Elvira, B. (2011). Ontogenetic variation in density-dependent growth of brown trout through habitat competition. *Freshwater Biology*, **56**(3), 530–540.

Pavey, S. A., Gaudin, J., Normandeau, E., Dionne, M., Castonguay, M., Audet, C., & Bernatchez, L. (2015). RAD sequencing highlights polygenic discrimination of habitat ecotypes in the panmictic American eel. *Current Biology*, **25**(12), 1666-71.

Pelgrom, S. M. G. J., Lock, R. A. C., Balm, P. H. M., & Bonga, S. E. W. (1995). Integrated physiological response of tilapia, *Oreochromis mossambicus*, to sublethal copper exposure. *Aquatic Toxicology*, **32**(4), 303–320.

Pentreath, R.J. (1994). The discharge of waters from active and abandoned mines. In: R.E. Hestor & R.M. Harrison (eds.) *Issues in Environmental Science and Technology I – Mining and its environmental impact*. Royal Society of Chemistry, Cambridge, UK. pp. 121-131.

Pérez-figueroa, A., García-pereira, M. J., Saura, M., Rolán-Alvarez, E., & Caballero, A. (2010). Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology*, **23**(10), 2267–2276.

Pickering, K.T. & Owen, L.A. (1997). *An introduction to global environmental issues*. London, New York: Routledge.

Pigliucci, M. & Schlichting, C. D., (1997). On the limits of quantitative genetics for the study of phenotypic evolution. *Acta Biotheoretica*, **45** (2)*:*143-160.

Pigliucci, M., (2006). Genetic variance-covariance matrices: a critique of the evolutionary genetics research program. *Biology & Philosophy*, **21**(1): 1-23.

Pinto, P. I., *et al.,* (2010). Gill transcriptome response to changes in environmental calcium in the green spotted puffer fish. *BMC Genomics*, **11**(1), 476.

Pirrie D, Beer AJ, Camm GS. (1999) Early diagenetic sulphide minerals in the Hayle Estuary, Cornwall, *Geoscience in South-West England*, **9**(4): 325-332.

Pirrie D, Power M, Rollinson GK, Camm G, Hughes S, Butcher A, Hughes P. (2003) The spatial distribution and source of arsenic, copper, tin and zinc within the surface sediments of the Fal Estuary, Cornwall, UK, *Sedimentology*, **50**: 579-595.

Pirrie D, Power MR, Payne A, Camm GS, Wheeler PD. (2000) Impact of mining on sedimentation; the Camel and Gannel estuaries, Cornwall, *Geoscience in South-West England*, **10**(1): 21-28.

Pirrie, D., Power, M. R., Wheeler, P. D., Cundy, A., Bridges, C., & Davey, G. (2002). Geochemical signature of historical mining: Fowey Estuary, Cornwall, UK. *Journal of Geochemical Exploration*, **76**(1), 31–43.

Pirrie, D., Sear, L. G., Hughes, S. H., & Camm, G. S. (1997). Mineralogical and geochemical signature of mine waste contamination, Tresillian River, Fal Estuary, Cornwall, UK. *Environmental Geology*, **29**(1–2), 58–65.

Plath, M., Pfenninger, M., Lerp, H., Riesch, R., Eschenbrenner, C., Slattery, P. A., Bierbach, D., Herrmann, N., Schulte, M., Arias–Rodriguez, L., Indy, J. R., Passow, C. & Tobler, M., (2013). Genetic differentiation and selection against migrants in evolutionarily replicated extreme environments. *Evolution*, **67**(9), 2647–2661.

Pollock, M. S., Clarke, L. M. J., & Dubé, M. G. (2007). The effects of hypoxia on fishes: from ecological relevance to physiological effects. *Environmental Reviews*, **15**: 1–14.

Polyakov, A., Beharav, A., Avivi, A. & Nevo, E., (2004). Mammalian microevolution in action: adaptive edaphic genomic divergence in blind subterranean mole-rats. *Proc. Biol. Sci.,* **271**(4): S156-S159.

Potts, D. (2002). *Project planning and analysis for development*. Boulder, Colorado: Lynne Reinner.

Pritchard, J. K., Pickrell, J. K., & Coop, G. (2010). The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology*, **20**(4), R208–R215.

Prodöhl, P. A., Taggart, J. B. and Ferguson, A. (1992), Genetic variability within and among sympatric brown trout (*Salmo trutta*) populations: multi-locus DNA fingerprint analysis. *Hereditas*, **117**: 45–50.

Prout, T., & Barker, J. S. (1993). F statistics in *Drosophila buzzatii*: selection, population size and inbreeding. *Genetics*, **134**(1), 369-375.

Putnam, H. M., Davidson, J. M., & Gates, R. D. (2016). Ocean acidification influences host DNA methylation and phenotypic plasticity in environmentally susceptible corals. *Evolutionary Applications*, **9**(9), 1165–1178.

Quinn, N. L., *et al.,* (2010). Genomic organization and evolution of the Atlantic salmon hemoglobin repertoire. *BMC Genomics,* **11**(1), 1099–1117.

Radford, A., & Gibson, H., (2004). Research report: The economic impact of game and coarse Angling in Scotland. Scottish Government Report.

Reid, S. D., McDonald, D. G., & Rhem, R. R. (1991). Acclimation to sublethal aluminum: modifications of metal – gill surface interactions of juvenile rainbow trout (*Oncorhynchus mykiss*). *Canadian Journal of Fisheries and Aquatic Sciences*, **48**(10), 1996–2005.

Reiser, D. W., & White, R. G. (1988). Effects of two sediment size-classes on survival of steelhead and chinook salmon eggs. *North American Journal of Fisheries Management*, **8**(4), 432–437.

Richards, P. M., Liu, M. M., Lowe, N., Davey, J. W., Blaxter, M. L., & Davison, A. (2013). RAD-Seq derived markers flank the shell colour and banding loci of the *Cepaea nemoralis* supergene. *Molecular Ecology*, **22**(11), 3077–3089.

Roda, F., *et al.,* (2013). Convergence and divergence during the adaptation to similar environments by an Australian groundsel. *Evolution*, **67**(9), 2515–2529.

Roesijadi, G. (1992). Metallothioneins in metal regulation and toxicity in aquatic animals. *Aquatic Toxicology*, **22**(2), 81–113.

Rogers, J., Richards, J., & Wood, C. (2003). Ionoregulatory disruption as the acute toxic mechanism for lead in the rainbow trout (*Oncorhynchus mykiss*). *Aquatic Toxicology*, **64**(2), 215–234.

Rollinson GK, Pirrie D, Power M, Cundy A, Camm G. (2008) Geochemical and Mineralogical Record of Historical Mining, Hayle Estuary, Cornwall, UK, *Geoscience in south-west England,* **11**:326-337.

Ryder, O. A. (1986). Species conservation and systematics: the dilemma of subspecies. *Trends in Ecology & Evolution*, **1**(1), 9–10.

Salway, P. (1993). A history of Roman Britain. Oxford: Oxford University Press.

Sanz Ball-llosera, N., Garcìa-Marìn, J. L., & Pla, C. (2002). Managing fish populations under mosaic relationships. The case of brown trout (*Salmo trutta*) in peripheral Mediterranean populations. *Conservation Genetics*, **3**(4): 385–400.

Sanz, N., Cortey, M., Pla, C., & García-Marín, J. L. (2006). Hatchery introgression blurs ancient hybridization between brown trout (*Salmo trutta*) lineages as indicated

by complementary allozymes and mtDNA markers. *Biological Conservation*, **130**(2), 278–289.

Sanz, N., García-Marín, J.-L., & Pla, C. (2000). Divergence of brown trout (*Salmo trutta*) within glacial refugia. *Canadian Journal of Fisheries and Aquatic Sciences*, **57**(11), 2201–2210.

Sarkar, A., Ray, D., Shrivastava, A. N., & Sarker, S. (2006). Molecular Biomarkers: Their significance and application in marine pollution monitoring. *Ecotoxicology*, **15**(4), 333–340.

Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics*, **14**(11), 807–820.

Schenekar, T., Lerceteau-Köhler, E., & Weiss, S. (2014). Fine-scale phylogeographic contact zone in Austrian brown trout *Salmo trutta* reveals multiple waves of post-glacial colonization and a pre-dominance of natural versus anthropogenic admixture. *Conservation Genetics*, **15**(3), 561–572.

Scheurer, K., Alewell, C., Bänninger, D., & Burkhardt-Holm, P. (2009). Climate and land-use changes affecting river sediment and brown trout in alpine countries—a review. *Environmental Science and Pollution Research*, **16**(2), 232–242.

Schlötterer, C. (2000). Evolutionary dynamics of microsatellite DNA. *Chromosoma*, **109**(6), 365–71.

Schlötterer, C. (2002). A Microsatellite-Based Multilocus Screen for the Identification of Local Selective Sweeps. *Genetics*, **160**(2).

Schlötterer, C. (2004). Opinion: The evolution of molecular markers — just a matter of fashion? *Nature Reviews Genetics*, **5**(1), 63–69.

Schlötterer, C., (2003). Hitchhiking mapping—functional genomics from the population genetics perspective. *Trends Genet*. **19**:32-38.

Scottish Environment Protection Agency. 2002. Managing river habitats for fisheries.

Scown, T. M., Santos, E. M., Johnston, B. D., Gaiser, B., Baalousha, M., Mitov, S., Lead, J.R., Stone, V., Fernandes, R., Jepson, M., van Aerle, R., & Tyler, C.R. (2010).

Effects of Aqueous Exposure to Silver Nanoparticles of Different Sizes in Rainbow Trout. *Toxicological Sciences*, **115**(2), 521–534.

Scrivener, R.C., Highley, D.E., Cameron, D.G., Linley, K.A., & White, R. (1997). Mineral resource information for development plans Phase One Cornwall: Resources and constraints. British Geological Society Technical Report WF/97/11 Mineral Resources Series.

Shapiro, B.,*et al*. (2004). Rise and fall of the Beringian Steppe Bison. *Science*, ***306***(5701), 1561-1565.

Shaw, R. G. & Etterson, J. R., 2012. Rapid climate change and the rate of adaptation: insight from experimental quantitative genetics. *New Phytologist*, **195**: 752–765.

Silva, A. O. F. da, & Martinez, C. B. R. (2014). Acute effects of cadmium on osmoregulation of the freshwater teleost Prochilodus lineatus: Enzymes activity and plasma ions. *Aquatic Toxicology*, **156**, 161–168.

Smedley, P. L. (1991). The geochemistry of rare earth elements in groundwater from the Carnmenellis area, southwest England. *Geochimica et Cosmochimica Acta*, **55**(10), 2767–2779.

Smith, S., *et al.,* (2013). RNA-seq analysis reveals extensive transcriptional plasticity to temperature stress in a freshwater fish species. *BMC Genomics*, **14**(1), 375.

Snell-Rood, E. C., Van Dyken, J. D., Cruickshank, T., Wade, M. J., & Moczek, A. P. (2010). Toward a population genetic framework of developmental evolution: the costs, limits, and consequences of phenotypic plasticity. *BioEssays*, 32(1), 71–81.

Sommer, S., *et al.,* (2005). The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Frontiers in Zoology*, 2(1), 16.

Southern, P. (2012). *Roman Britain: A New History 55 BC-AD 450*. Gloucestershire: Amberly Publishing

Spitze, K. (1993). Population structure in *Daphnia obtusa*: quantitative genetic and allozymic variation. *Genetics*, **135**(2), 367-374.

Stabell, O. B. (1984). Homing andolfaction in Salmonids: A critical review with special reference to the Atlantic Salmon. *Biological Reviews*, **59**(3), 333–388.

Stanhope, M.J., Hartwick, B., & Baillie, D., 1993. Molecular phylogeographic evidence for multiple shifts in habitat preference in the diversification of an amphipod species. Molecular Ecology, **2**(2)**:** 99–112.

Stanier, P. (2011). *Cornwall's Geological Heritage*. Cornwall, UK: Twelveheads Press.

Stapley, J., *et al.,* (2010). Adaptation genomics: the next generation. *Trends in Ecology & Evolution*, **25**(12), 705–712.

Stapley, J.*,* Reger, J.*,* Feulner, P.G., Smadja, C., Galindo, J., Ekblom, R., Bennison, C., Ball, A.D., Beckerman, A.P., Slate, J., 2010. Adaptation genomics: the next generation*.* Trends in Ecology and Evolution*,* **25**(12): 705–712*.*

Stelkens, R. B., Pompini, M., & Wedekind, C. (2012). Testing for local adaptation in brown trout using reciprocal transplants. *BMC Evolutionary Biology*, **12**(247). DOI: 10.1186/1471-2148-12-247.

Stephan, W. (2016). Signatures of positive selection: from selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. *Molecular Ecology*, **25**(1), 79–88.

Stevens, J. R., *et al.,* (1992). Numerical taxonomy of Trypanozoon based on polymorphisms in a reduced range of enzymes. *Parasitology*, **104**(1), 75.

Stuart, T. (1957). The migrations and homing behaviour of brown trout. *Freshw. Salmon Fish. Res.* **18**: 305-316.

Suárez, J., Bautista, J. M., Almodóvar, A., & Machordom, A. (2001). Evolution of the mitochondrial control region in Palaearctic brown trout (*Salmo trutta*) populations: the biogeographical role of the Iberian Peninsula. *Heredity*, **87**(2): 198–206.

Sudmant, P. H., *et al.,*(2015). Global diversity, population stratification, and selection of human copy-number variation. *Science*, **349**(6253), aab3761.

Suttle, K. B., Power, M. E., Levine, J. M., & McNeely, C. (2004). How fine sediment in riverbeds impairs growth and survival of juvenile salmonids. *Ecological Applications*, **14**(4), 969–974.

Taylor, E. B. (1991). A review of local adaptation in Salmonidae, with particular reference to Pacific and Atlantic salmon. *Aquaculture*, **98**(1-3), 185–207.

Taylor, L. N., Wood, C. M., & McDonald, D. G. (2003). An evaluation of sodium loss and gill metal binding properties in rainbow trout and yellow perch to explain species differences in copper tolerance. *Environmental Toxicology and Chemistry*, **22**(9), 2159.

Thorndycraft, V. R., Pirrie, D., & Brown, A. G. (2004). Alluvial records of medieval and prehistoric tin mining on Dartmoor, southwest England. *Geoarchaeology*, **19**(3), 219–236.

Thorstad, E. B., Økland, F., Aarestrup, K., & Heggberget, T. G. (2008). Factors affecting the within-river spawning migration of Atlantic salmon, with emphasis on human impacts. *Reviews in Fish Biology and Fisheries*, **18**(4), 345–371.

Tierney, K. B., Baldwin, D. H., Hara, T. J., Ross, P. S., Scholz, N. L., & Kennedy, C. J. (2010). Olfactory toxicity in fishes. *Aquatic Toxicology*, **96**(1), 2–26.

Timberlake, S., 2001. Mining and prospection for metals in early Bronze Age Britain – making claims within the archeological landscape. In J. Bruck (ed.). *Bronze Age Landscapes: tradition and transformation*, Oxford, Oxbow, pp. 179-92.

Tom, M., Chen, N., Segev, M., Herut, B., & Rinkevich, B. (2004). Quantifying fish metallothionein transcript by real time PCR for its utilization as an environmental biomarker. *Marine Pollution Bulletin*, **48**(7), 705–710.

Townsend, C. R. & Crowl, T. A. (1991). Fragmented population structure in a native New Zealand fish: an effect of introduced brown trout? *Oikos*, **61**(3), 347-354.

Trynka, G., *et al.,* (2011). Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nature Genetics*, **43**(12), 1193–1201.

Trynka, G., Hunt, K. A., Bockett, N. A., Romanos, J., Mistry, V., Szperl, A., *et al.* van Heel, D. A. (2011). Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nature Genetics*, **43**(12), 1193–1201.

Turner, A. (2000). Trace Metal Contamination in Sediments from U.K. Estuaries: An Empirical Evaluation of the Role of Hydrous Iron and Manganese Oxides. *Estuarine, Coastal and Shelf Science*, **50**(3), 355–371.

Turner, T.L., von Wettberg, E.J. & Nuzhdin, S.V., 2008. Genomic analysis of differentiation between soil types reveals candidate genes for local adaptation in *Arabidopsis lyrata*. PLoS ONE **3**(9): e3183.

Ugolini, A., Borghini, F., Calosi, P., Bazzicalupo, M., Chelazzi, G. & Focardi, S., 2004. Mediterranean *Talitrus saltator* (*Crustacea, Amphipoda*) as a biomonitor of heavy metals contamination. Marine Pollution Bulletin, **48**(5-6): 526-532.

Uren Webster, T. M., & Santos, E. M. (2015). Global transcriptomic profiling demonstrates induction of oxidative stress and of compensatory cellular stress responses in brown trout exposed to glyphosate and Roundup. *BMC Genomics*, **16**(1), 32.

Uren Webster, T. M., Perry, M. H., & Santos, E. M. (2015). The Herbicide Linuron Inhibits Cholesterol Biosynthesis and Induces Cellular Stress Responses in Brown Trout. *Environmental Science & Technology*, **49**(5), 3110–3118.

Valladares, F., Gianoli, E., & Gómez, J. M. (2007). Ecological limits to plant phenotypic plasticity. *New Phytologist*, **176**(4), 749–763.

van Aerle, R., Lange, A., Moorhouse, A., Paszkiewicz, K., Ball, K., Johnston, B. D., De-Bastos, E., Booth, T., Tyler, C.R., & Santos, E.M. (2013). Molecular mechanisms of toxicity of silver nanoparticles in zebrafisheEmbryos. *Environmental Science & Technology*, **47**(14), 8005–8014.

van Kleunen, M., & Fischer, M. (2005). Constraints on the evolution of adaptive phenotypic plasticity in plants. *New Phytologist*, **166**(1), 49–60.

van Straalen, N. M. & Timmermans, M.J.T.N., 2002. Genetic variation in toxicant-stressed populations: An evaluation of the "genetic erosion" hypothesis. Human and Ecological Risk Assessment **8**: 983-1002.

Verspoor, E. (1988). Widespread hybridization between native Atlantic salmon, *Salmo salar*, and introduced brown trout, *S. trutta*, in eastern Newfoundland. *Journal of Fish Biology*, **32**(3), 327–334.

Vignal, A., Milan, D., SanCristobal, M. & Eggen, A., 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Genet. Sel. Evol., **34**:275-305.

Vitalis, R., Dawson, K., Boursot, P., & Belkhir, K. (n.d.). DetSel 1.0: a computer program to detect markers responding to selection. *The Journal of Heredity*, **94**(5), 429–31.

Vuorinen, J., & Piironen, J. (1984). Electrophoretic Identification of Atlantic Salmon (*Salmo salar*), Brown Trout (*S . trutta*), and Their Hybrids. *Canadian Journal of Fisheries and Aquatic Sciences*, **41**(12), 1834–1837.

Walling, D. E., Collins, A. L., & McMellin, G. K. (2003). A reconnaissance survey of the source of interstitial fine sediment recovered from salmonid spawning gravels in England and Wales. *Hydrobiologia*, **497**(1/3), 91–108.

Walsh, M. R., *et al.,* (2016). Local adaptation in transgenerational responses to predators. *Proceedings of the Royal Society of London B: Biological Sciences*, **283**(1823). DOI: 10.1098/rspb.2015.2271

Wan, Q.-H., Wu, H., Fujihara, T. & Fang, S.-G., 2004. Which genetic marker for which conservation genetics issue? *Electrophoresis*, **25**(14): 2165–2176.

Wang, Z., Gerstein, M. & Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nature Rev. Genetics **10**(1): 57-63.

Waples, R. (1995). Evolutionarily significant units and the conservation of biological diversity under the Endangered Species Act. *American Fisheries Society Symposium*

Watt, W. B. (1994). Allozymes in evolutionary genetics: self-imposed burden or extraordinary tool? *Genetics*, **136**(1), 11-16.

Watt, W.B., Wheat, C.W., Meyer, E.H., Martin, J.F., 2003. Adaptation at specific loci. VII. Natural selection, dispersal and the diversity of molecular-functional variation patterns among butterfly species complexes (*Colias: Lepidoptera, Pieridae*). Molecular Ecology, **12**(5): 1265-75.

Weiss, S., Antunes, A., Schlotterer, C., & Alexandrino, P. (2000). Mitochondrial haplotype diversity among Portuguese brown trout *Salmo trutta L*. populations: relevance to the post-Pleistocene recolonization of northern Europe. *Molecular Ecology*, **9**(6), 691–698.

Weldon, W. F. R., 1895. Attempt to measure the death-rate due to the selective destruction of *Carcinus moenas* with respect to a particular dimension. *Proc. R. Soc. Lond.* **58:** 360–379.

Wells, A., *et al,*. (2006). Physiological effects of simultaneous, abrupt seawater entry and sea lice (*Lepeophtheirus salmonis*) infestation of wild, sea-run brown trout (*Salmo trutta*) smolts. *Canadian Journal of Fisheries and Aquatic Sciences*, **63**(12), 2809–2821.

Went, A. E. J. (1979). 'Ferox' trout, *Salmo trutta L*. of Loughs Mask and Corrib. *Journal of Fish Biology*, **15**(3), 255–262.

West, S., Charman, D. J., Grattan, J. P., & Cherburkin, A. K. (1997). Heavy metals in Holocene peats from south west England: Detecting mining impacts and atmospheric pollution. *Water, Air, and Soil Pollution*, **100**(3/4), 343–353.

Wheat, C.W., Watt, W.B., Pollock, D.D. & Schulte, P.M., 2006. From DNA to fitness differences: sequences and structures of adaptive variants of *Colias* Phosphoglucose Isomerase (PGI). *Molecular Ecology*, **23**(3): 499-512.

Whitlock, M. C., & Guillaume, F. (2009). Testing for spatially divergent selection: Comparing Q_{ST} to F_{ST}. *Genetics*, **183**(3), 1055-1063.

Whitman, D.W., & Agrawal, A.A., 2009. "What is phenotypic plasticity and why is it important", In: D.W. Whitman, T.N. Ananthakrishnan (eds). *Phenotypic Plasticity of Insects: Mechanisms and Consequences*, Enfield, USA: Science Publishers, pp. 1-63.

Whittaker, J.C., Harbord, R.M., Boxall, N., Mackay, I., Dawson, G. & Silby, R.M., 2003. Likelihood-based estimation of microsatellite mutation rates. Genetics, **164**(2): 781-787.

Wickett, W.P. (1954). The oxygen supply to salmon eggs in spawning beds. *Journal of the Fisheries Research Board of Canada*, **11**(6): 933-953.

Wilding, C. S., Butlin, R. K. & Grahame, J., 2001. Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *Journal of Evolutionary Biology*, **14**(4): 611–619.

Wood, C. M. (1992). Flux measurements as indices of H^+ and metal effects on freshwater fish. *Aquatic Toxicology*, **22**(4), 239–263.

Wood, C. M., Farrell, A. P., & Brauner, C. J. (2012). *Homeostasis and toxicology of essential metals*. Academic Press.

Wright, K. M., *et al*. (2013). Indirect Evolution of Hybrid Lethality Due to Linkage with Selected Locus in Mimulus guttatus. *PLoS Biology*, **11**(2), e1001497.

Wu, S.-M., Tsai, J.-W., Tzeng, W.-N., Chen, W.-Y., & Shih, W.-Y. (2015). Analyzing the effectiveness of using branchial NKA activity as a biomarker for assessing waterborne copper toxicity in tilapia (*Oreochromis mossambicus*): A damage-based modeling approach. *Aquatic Toxicology*, ***163*,** 51–59.

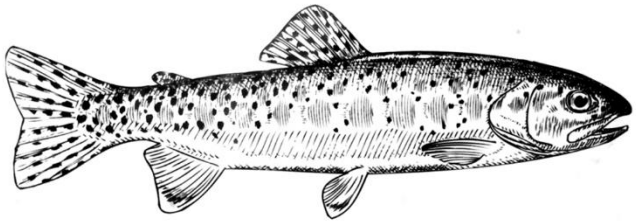
Yeaman, S., *et al.,* (2016). Convergent local adaptation to climate in distantly related conifers. *Science*, **353**(6306), 1431-1433.

Yim, W. W.-S. (1981). Geochemical investigations on fluvial sediments contaminated by tin-mine tailings, Cornwall, England. *Environmental Geology*, **3**(5), 245–256.

Yim, W.W.S. (1976). Heavy metal accumulation in estuarine sediments in a historical mining of Cornwall. *Marine Pollution Bulletin* **7**:147-150.

Zimmer, A. M., Barcarolli, I. F., Wood, C. M., & Bianchini, A. (2012). Waterborne copper exposure inhibits ammonia excretion and branchial carbonic anhydrase activity in euryhaline guppies acclimated to both fresh water and sea water. *Aquatic Toxicology*, **122**, 172–180.

Zueva, K. J., *et al.,* (2014). Footprints of directional selection in wild Atlantic salmon populations: Evidence for parasite-driven evolution? *PLoS ONE*, **9**(3), e91672.



Chapter II

Human mining activity across the ages determines the genetic structure of modern brown trout (*Salmo trutta*) populations

Published in *Evolutionary Applications* (2015) **8**(6) 573-585

AUTHORS: J.R. Paris[1], R.A King[1] and J.R. Stevens[1]

[1] Biosciences, College of Life and Environmental Sciences, University of Exeter, Exeter, UK

J.R. Paris assisted in designing the research, conducted the laboratory work, data analyses and wrote the manuscript

R.A. King assisted in designing the research, conducted laboratory work, assisted with data analysis and improved the manuscript

J.R. Stevens designed the research and improved the manuscript

Evolutionary Applications

Open Access

Evolutionary Applications ISSN 1752-4571

ORIGINAL ARTICLE

Human mining activity across the ages determines the genetic structure of modern brown trout (*Salmo trutta* L.) populations

Josephine R. Paris, R. Andrew King and Jamie R. Stevens

Biosciences, College of Life and Environmental Sciences, University of Exeter, Exeter, UK

Keywords
anthropogenic, DIYABC, genetic diversity, metal contamination, microsatellite, mining activity, population structure.

Correspondence
Jamie R. Stevens, Biosciences, College of Life and Environmental Sciences, University of Exeter, Geoffrey Pope Building, Stocker Road, Exeter EX4 4QD, UK.
Tel.: +44-1392263775;
fax: +44-1392263700;
e-mail: J.R.Stevens@exeter.ac.uk

Received: 23 January 2015
Accepted: 13 April 2015

doi:10.1111/eva.12266

Abstract

Humans have exploited the earth's metal resources for thousands of years leaving behind a legacy of toxic metal contamination and poor water quality. The southwest of England provides a well-defined example, with a rich history of metal mining dating to the Bronze Age. Mine water washout continues to negatively impact water quality across the region where brown trout (*Salmo trutta* L.) populations exist in both metal-impacted and relatively clean rivers. We used microsatellites to assess the genetic impact of mining practices on trout populations in this region. Our analyses demonstrated that metal-impacted trout populations have low genetic diversity and have experienced severe population declines. Metal-river trout populations are genetically distinct from clean-river populations, and also from one another, despite being geographically proximate. Using approximate Bayesian computation (ABC), we dated the origins of these genetic patterns to periods of intensive mining activity. The historical split of contemporary metal-impacted populations from clean-river fish dated to the Medieval period. Moreover, we observed two distinct genetic populations of trout within a single catchment and dated their divergence to the Industrial Revolution. Our investigation thus provides an evaluation of contemporary population genetics in showing how human-altered landscapes can change the genetic makeup of a species.

Introduction

The exploration for and exploitation of metals have played a crucial role in human history. Although metals are natural constituents of the earth's crust (Wedepohl 1991), their prevalence within natural systems has been significantly enhanced by human activity (Han et al. 2002). On a global basis, approximately half of all metal fluxes in the environment are anthropogenically driven (Wood 2011). Metals are extremely persistent in the environment; they are non-degradable, and thus, readily accumulate at toxic levels. Mining for such metals has a rich history in Britain. In particular, large areas of southwest England have been mined since the Bronze Age (2500 BCE: Dines 1956; Buckley 2002), with increasing exploitation during the Roman occupation (McFarlane et al. 2013), and later, as technology improved, throughout the Medieval period (Lewis 1965), with activity reaching its peak in the 19th century during the Industrial Revolution (Rainbow et al. 2011) when the region became the world's leading producer of many economically important metals (Dines 1956; Penhallurick 1986).

Natural populations are predicted to experience changes in genetic diversity and population structure, especially through genetic drift, gene flow and/or selection, but these genetic shifts have been shown to be amplified by anthropogenic interference (Banks et al. 2013)—for example, through habitat loss and fragmentation (Mondol et al. 2013); hunting pressure (González-Porter et al. 2011); overfishing (Allendorf et al. 2014); invasive species (Austin et al. 2011); and disease transmission (Kyle et al. 2014). Genetic data can provide a beneficial insight into this context. With respect to neutral markers, signatures of population-level disturbance may include reduced genetic

diversity and associated population bottlenecks (Fontaine et al. 2012), alterations in population structure and disruption of gene flow (Clark et al. 2010), and signatures of loci under selection (Lind and Grahn 2011). The ability to quantify and understand these genetic processes is essential to informing conservation and management practices (Amos and Balmford 2001; Weeks et al. 2011).

Brown trout (*Salmo trutta* L.) populations are known to reside in rivers across southwest England, in both so-called clean and metal-impacted rivers. Due to the underlying geology (Webb et al. 1978) and ancient history of mining (Buckley 2002), metals are present in such 'clean' rivers, but at relatively low concentrations: River Camel (total zinc ~17 μg/L; total copper ~5 μg/L; total arsenic ~4 μg/L); River Fal (total zinc ~37 μg/L; total copper ~5 μg/L; total arsenic ~4 μg/L) (Environment Agency Data). On the other hand, metal rivers are defined by being heavily impacted by known historical mining activity. These rivers contain significantly elevated concentrations of metals: River Hayle (total zinc ~350 μg/L; total copper ~28 μg/L; total arsenic ~9 μg/L); Red River (total zinc ~238 μg/L; total copper ~27 μg/L; total arsenic ~86 μg/L) (Environment Agency Data). As these rivers are known to vary in their contamination levels, gradient effects between metal contaminant exposure and variation in genetic patterns of brown trout can be assessed.

Metal loads within such rivers can vary spatially over just a few kilometers (Environment Agency Data) meaning that genetic substructure may exist not only between, but also within a single system (Vähä et al. 2007). The River Hayle is a particularly well-studied metal-contaminated catchment. The whole river is affected by toxic metal pollution, dating back to the Industrial Revolution (19th Century), but is punctuated by an extremely high metal-contaminated middle region at Godolphin Cross (total zinc ~2512 μg/L; total copper ~417 μg/L; total arsenic ~99 μg/L: Environment Agency Data). Metal contamination of such high levels is known to cause acute toxicity in metal-naïve brown trout (Hansen et al. 2006a,b; Durrant et al. 2011), yet brown trout exist along the river, both upstream and downstream of this middle region.

Using a panel of microsatellites, we sought to establish the patterns of genetic diversity and genetic structuring of trout from metal-impacted populations compared to trout from clean control rivers. Specifically, (i) Is genetic diversity reduced in metal-impacted populations?; (ii) Can these patterns be explained by genetic evidence of population bottlenecks?; (iii) Do metal-impacted populations show distinct genetic structuring due to reduced gene flow and genetic drift?; (iv) Given the long history of mining in the region, what is the most likely genetic origin of contemporary metal populations?; and finally, (v) Can the genetic changes we observe be linked to periods of increased mining activity? With a potential tie to human-mediated evolutionary change, analysis characterizing a temporal component to these genetic changes in a historical context is vital (Smith and Bernatchez 2008). To fully explore the spatial extent of mining practices on trout populations, we have taken a multiscaled approach, exploring evidence of genetic impacts across a larger region (southwest England) as well as conducting a microgeographic analysis of the River Hayle.

Materials and methods

Populations and sampling

A total of 700 individual brown trout (*Salmo trutta* L.) were included for study (Fig. 1, Table S1). To account for within-river variation in metal-contamination levels, where possible, two sites per river were sampled. Of the 15 populations, six were selected as 'clean' control populations and the other nine were classified as originating from mining-impacted sites (Table S1). Brown trout sampled from each of these geographic sites will be referred to, and treated as populations. To the author's knowledge, there is no known history of supplementary stocking on the rivers included in this study.

Fish (1 + parr or older) were sampled as part of routine Environment Agency monitoring programs; briefly, fish were anesthetized using either clove oil or benzocaine (10 g/100 mL ethanol) diluted 1:2000 in river water prior to adipose fin clip removal. Fin clips were placed into 95% ethanol for storage.

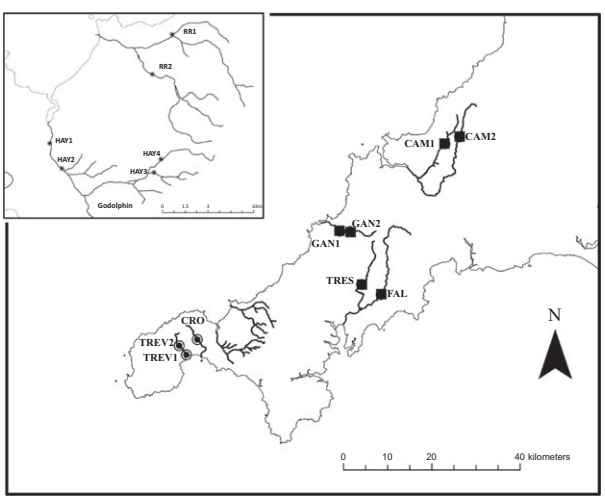


Figure 1 Geographic location of the populations sampled. Site codes correspond to those given in Table S1. Black squares represent clean sites, and black circles represent metal-contaminated sites. The enlarged map shows the sites on the Red River and the River Hayle.

Microsatellite loci selection

Twenty-two putative neutral microsatellite loci were screened: Ssa85 and Ssa197 (O'Reilly et al. 1996); Ssa52NVH (Gharbi et al. 2006); SsosL417 and SsosL311 (Slettan et al. 1995); SS11 (Martinez et al. 1999); CA048828, BG935488, CA060208, CA060177, CA053293, CA515794 and CA769358 (Vasemägi et al. 2005); SsaD157 and SsaD58 (King et al. 2005); SsaF43 (Sánchez et al. 1996); Str3QUB (Keenan et al. 2013); Ssa407UOS and Ssa412UOS (Cairney et al. 2000); SSsp2213 (Paterson et al. 2004); One102 (Olsen et al. 2000 - using the primers of Keenan et al. 2013). Three MHC class 1-linked loci (sasaTAP2A, sasaTAP2B and sasa-UBA: Grimholt et al. 2002) were also chosen for screening due to their association with regions potentially under selection (Zelikoff 1993; Cohen 2002).

DNA extraction, PCR, and genotyping

DNA was extracted from adipose fin clips using the HotSHOT method (Truett et al. 2000). PCR amplification was carried out using a BIO-RAD MyCycler Thermal Cycler in 10 μL reaction volumes containing 1 μL of extracted DNA (c. 30 ng DNA), 3 μL RNase-free water, 5 μL of QIAGEN HotStarTaq *Plus* Master Mix, and 1 μL of primer mixture, in a total of 8 microsatellite multiplexes (Table S2). PCR conditions were as follows: an initial denaturing step at 95°C for 5 min was followed by 20 cycles of touchdown PCR consisting of 30 s at 94°C, a 30 s annealing step starting at 60°C or 55°C and decreasing by 0.5°C each cycle until a touchdown temperature of 50°C or 45°C (dependent on multiplex; Table S2) was achieved, and an elongation step of 72°C for 30 s, followed by 15 cycles comprising 94°C for 30 s, 50°C or 45°C for 30 s, and 72°C for 30 s. This was followed by a final 10 min extension step at 72°C. Genotyping was performed on a Beckman Coulter CEQ™ 8000 Genetic Analysis System.

Full-sib analysis

To prevent biasing estimates of population genetic parameters, full-sibs within each population were identified using COLONY v 2.0.4.0 (Jones and Wang 2010). Each population was analyzed separately using a full-likelihood (FL) method, with a high-likelihood precision and a medium-length run. Two runs were performed under different random seeds. Any observed inconsistencies between these runs resulted in replicate runs being undertaken until the results were in consensus. Fish were considered members of a full-sib family if the probability of exclusion as full-sib families was >0.9. A single individual of each full-sib family was retained in the dataset.

Data quality

The occurrence of homozygote excess, stuttering, large allele dropout, and null alleles were assessed using MICROCHECKER v 2.2 (van Oosterhout et al. 2004). Tests for linkage disequilibrium (LD) and Hardy–Weinberg equilibrium (HWE) were performed using GENEPOP v 4.2 (Rousset 2008), implementing the log likelihood ratio statistic and probability test, respectively. Default Markov chain parameters were used for both analyses. False discovery rate (FDR) was used to detect Type I errors in tests for LD and HWE (Storey and Tibshirani 2003).

Genetic diversity

Three measures of genetic diversity were determined for each population. Allelic richness (A_R) was calculated using FSTAT 2.9.3 (Goudet 1995) based on a minimum sample size of 26 diploid individuals. Observed heterozygosity (H_O) and expected heterozygosity (H_E) were calculated using GENALEX 6.5 (Peakall and Smouse 2012). Statistical differences in genetic diversity between population groups were calculated using FSTAT 2.9.3 (Goudet 1995), based on 1000 permutations of the dataset.

Population bottlenecks

We used two approaches to test whether mining activity in the region has caused demographic declines in the studied trout populations. The program BOTTLENECK (Cornuet and Luikart 1996) uses the degree of heterozygosity excess compared to expectations under mutation–drift equilibrium to quantify the loss of rare alleles shortly after bottlenecks. Peery et al. (2012) have suggested that estimation of the number of multistep mutations at microsatellite loci is often underestimated and thus recommended a stepwise mutation rate of $p_S = 0.78$ (Peery et al. 2012). We therefore implemented the two-phase model (TPM) with 80% stepwise mutations (SMM). Deviations between observed and expected frequency distributions were statistically tested using Wilcoxon's signed-rank test for one-tail heterozygosity excess ($P < 0.05$). BOTTLENECK was run for 10 000 iterations. To further explore evidence of population bottlenecks, we also calculated the *M*-ratio, which is the ratio of the number of alleles to the range in allele size. This is expected to be lower in bottlenecked populations due to the loss of rare alleles (Garza and Williamson 2001). We used a prebottleneck effective population size (N_E) between 50 and 100; theta (θ) thus varied from 0.1 to 0.2. We used a mutation rate of $\mu = 5 \times 10^{-4}$ and used $p_g = 0.22$ ($p_S = 0.78$) and $\delta_g = 3.1$, as suggested by Peery et al. (2012).

Detecting loci under selection

We used three tests to detect whether any of the microsatellite loci showed evidence of selection. We used Bayescan v 2.1 (Foll and Gaggiotti 2008), which extends the distance-based method of Beaumont and Balding (2004) by adopting a Bayesian hierarchical model, implemented via reversible jump Markov chain Monte Carlo (RJ-MCMC) simulations. We used a burn-in of 50 000 iterations and a thinning interval of 10. Sample size was set at 5000, resulting in a total of 10^5 iterations with 20 pilot runs (each with a length of 5000). We also used the statistic *lnRV* (Schlotterer 2002; Schlötterer and Dieringer 2005) to identify loci potentially hitchhiking with regions of the genome experiencing a selective sweep. We used the Bayesian two-way heterogeneous analysis of variance model (Marshall and Weiss 2006). Analyses were performed using WinBUGS v 1.4 (Lunn et al. 2000), implementing 20 000 MCMC iterations, with a 4000-iteration burn-in. Finally, we performed the Fdist test (Excoffier et al. 2009) in Arlequin v 3.5, which performs F_{ST} outlier simulations (Excoffier and Lischer 2010). We performed 3 separate runs of each dataset, each with 50 000 simulations under the hierarchical island model among groups of populations (F_{CT}) with 100 demes simulated per group and 10 simulated groups. We considered a locus to be under selection if the *P*-value was <0.01.

Population structure

Genetic differentiation was analyzed using global and pairwise F_{ST} estimates, calculated in GENALEX 6.5 (Peakall and Smouse 2012). Significance of estimates was based on 999 permutations of the dataset. F_{ST} was calculated for all loci in all populations, as well as independently for the River Hayle.

Genetic structuring of populations was analyzed using the model-based algorithm implemented in STRUCTURE v 2.3.3 (Pritchard et al. 2000), using a burn-in period of 50 000 iterations followed by 150 000 iterations with the number of inferred populations (*K*) ranging from 1 to 20. Ten independent runs of the program were performed using the population admixture model and correlated allele frequencies. The most likely number of population clusters was determined using the ΔK statistic (Evanno et al. 2005). To identify finer levels of structure, subsequent hierarchical analyses were performed based on the optimum *K* value from the first runs (Vähä et al. 2007). Analysis parameters for the hierarchical analyses were as given above except that maximum *K* was set at $n + 2$, where n represents the number of populations in the analysis.

To further explore population structure, a neighbor-joining dendrogram was constructed using the program POPULATIONS v1.2.28 (available at bioinformatics.org/~tryphon/populations), based on the Cavalli-Sforza and Edwards chord distance, D_{CE} (Cavalli-Sforza and Edwards 1967). Statistical significance of branches was tested by bootstrap analysis based on 1000 iterations. The dendrogram was visualized in MEGA v 6 (Tamura et al. 2013).

Placing time points on genetic divergence

We used approximate Bayesian computation (ABC) as implemented in the program DIYABC (Cornuet et al. 2008, 2014) to explore the genetic history of the contemporary trout populations, divergence times between these populations, and also to determine changes in effective population size (N_E). DIYABC outputs averages for each event in generations. We converted generations into dates using a generation time for brown trout of 4 years (Jensen et al. 2008).

Preliminary investigations

Based on the results of the population structure analyses of this study, we conducted preliminary investigations based on simple scenarios comparing the general clean group to each metal group: (i) Clean & Hayle, (ii) Clean & Red River, and (iii) Clean & Crowlas/Trevaylor. For the River Hayle analyses, we specified two time (t)-splitting points, t1: the split within the River Hayle and t2: the split of the Hayle populations from the clean populations. Conditions were set so that t2 > t1. For the Red River and Crowlas/Trevaylor investigations, only one time-splitting point was specified (t1), defining the split between the metal-impacted group and the clean group. For each of these independent population comparisons, we passed two scenarios, each composed of a reference table consisting of 10^5 simulated datasets. Scenario 2 differed from Scenario 1 by the inclusion of a reduction in population size following the split between the clean and metal populations, as suggested by genetic diversity estimates calculated here (see Results). Default minimum and maximum priors (10–10 000) were used for all parameters (N—effective population size; Nb—prebottleneck effective population size; t—time-splitting point(s); db—duration of bottleneck).

Hypotheses-testing scenarios

Based on the outcome of the preliminary investigations, we constructed more complicated scenarios that included genotypes from all of the populations, organized into 5 population groups: (i) clean populations; (ii) RR1&2 (Red River); (iii) CRO/TREV1&2 (Crowlas and Trevaylor); (iv) HAY1&2 (downstream Hayle); and (v) HAY3&4 (upstream Hayle). We constructed 3 groups of hypotheses-testing scenarios, the details of the topology of these can be found in Fig. S1. Group 1 consisted of scenarios by which the clean and metal populations are independently derived from a

common ancestor that is neither definitely clean nor metal in its genetic background. Group 2 scenarios tested variations on the hypothesis that metal populations are derived multiple times from a clean lineage. Group 3 comprised scenarios where the clean and metal groups are separate lineages, with metal populations being derived from the common metal lineage. Based on observations of preliminary runs, the maximum prior for the t1 parameter (Hayle split) was set at 3000 generations (12 000 years). The duration of bottleneck (db) was set at a maximum of 300 generations (1200 years). All other parameter priors remained as default values (10–10 000). Conditions were placed on splitting time points so that t4 > t3; t3 > t2; and t2 > t1.

Summary statistics and model checks

For all investigations and scenarios, we used the summary statistics of Cornuet et al. (2008): One-sample summary statistics included mean number of alleles; mean genic diversity; mean Garza-Williamson's *M,* and two-sample summary statistics included F_{ST} and mean classification index. We used the default priors for the mutation model (Min: 10^{-4}; Max: 10^{-3}; Mean: 5×10^{-4}). We performed model checking using a PCoA, and posterior probabilities of each scenario were then calculated using a logistic regression of 1% of the simulated data. After bias and precision estimations, model checks were performed for each scenario, using the summary statistics of Guillemaud et al. (2010), which include the two population summary statistics: mean number of alleles, mean genic diversity, mean size variance, $F_{ST.}$ and the classification index. For the group-based hypotheses scenarios, we simulated 10^5 runs for each scenario, after which we selected the scenario(s) from each group with the highest posterior probability. These scenarios were then compared in a final run, simulating 10^6 datasets per scenario, using the same priors, summary statistics, and model checks as outlined above.

Results

Data quality and correction

A total of 700 individuals from 15 brown trout populations were genotyped. Fifty-nine individuals (~8.4%) were removed due to sibling effects within the sampling. The 24 primer sets amplified a total of 25 loci, with the primers for One102 amplifying two loci with nonoverlapping size ranges (designated One102a and One102b). There were potential genotyping errors at 14 loci. We removed locus CA053293, as these inaccuracies occurred in 7 of the 15 populations. Evidence of homozygote excess and null alleles was not consistently detected in any other loci.

Indication of linkage disequilibrium (LD) between sasaUBA & sasaTAP2B was statistically significant in 12 of the 15 populations. There is evidence that these loci are in fact physically linked (Grimholt et al. 2002). Due to difficulty in scoring, it was decided to omit locus sasa-UBA from further analysis. Tests for Hardy–Weinberg equilibrium (HWE) revealed that five loci (Ssa412UOS, Ssa407UOS, Ssa52NVH, SsaD157, and SsosL417) showed significant deviation from HWE. However, as no locus was considered to be out of HWE in more than one population, all loci were retained in subsequent analyses.

Genetic diversity

Metal-impacted populations had lower genetic diversity in comparison with populations from the clean sites (Table 1). These differences were statistically significant (A_R, $P = 0.001$; H_E, $P = 0.001$; H_O, $P = 0.002$). Allelic richness (A_R) varied between 6.2 (HAY1) and 11.15 (CAM2). Similar patterns were observed for measures of heterozygosity: Expected heterozygosity (H_E) varied from 0.61 (HAY2) to 0.78 (CAM2 & FAL) and observed heterozygosity (H_O) ranged between 0.61 (HAY2) and 0.79 (CAM2). In particular, genetic diversity measures showed marked differences between the River Hayle and the River Camel, the mouths of which both flow out of north Cornwall; approximately 55 km from one another.

Across the metal populations, TREV1 had unusually high levels of genetic diversity, more similar to levels observed in clean populations. The River Hayle had very low genetic diversity when compared not only to the clean populations ($A_R = 0.005$; $H_E = 0.001$; $H_O = 0.001$) but also to other

Table 1. Measures of population genetic parameters for each population using 23 microsatellite loci.

Population	N_1	N_2	A_R	H_E	H_O	Wilcoxon TPM	M-ratio
CAM1	49	47	9.89	0.77	0.75	ns	0.59
CAM2	44	44	11.15	0.78	0.79	0.003	0.61
GAN1	50	49	9.51	0.75	0.76	ns	0.56
GAN2	50	45	8.92	0.74	0.73	0.052	0.58
FAL	47	42	10.25	0.78	0.77	ns	0.61
TRES	48	46	10.01	0.76	0.75	ns	0.60
RR1	45	41	6.87	0.70	0.70	0.000	0.50
RR2	41	40	7.79	0.70	0.69	0.008	0.57
HAY1	44	43	6.20	0.62	0.62	0.015	0.47
HAY2	48	39	6.23	0.61	0.61	0.006	0.49
HAY3	48	42	6.68	0.65	0.65	0.019	0.50
HAY4	37	27	6.34	0.63	0.64	ns	0.50
CRO	49	46	6.96	0.65	0.68	0.011	0.51
TREV1	50	45	9.36	0.74	0.74	ns	0.58
TREV2	50	45	7.30	0.70	0.70	ns	0.52

N_1—number of sampled individuals, N_2—number of individuals after full-sib removal, A_R —allelic richness, H_E—expected heterozygosity, H_O—observed heterozygosity, (i) Wilcoxon one-tail test results from BOTTLENECK, (ii) average *M* from *M*-ratio.

metal-impacted populations (A_R = 0.005; H_E = 0.08, H_O = 0.04). Sites downstream of the Godolphin region (HAY1, HAY2) had lower genetic diversity, compared to the two sites upstream (HAY3, HAY4), although these differences were not statistically significant. In particular, there was a marked difference in genetic diversity between the two sites straddling the metal region, which are separated by just 8 km; the upstream site (HAY3) had the highest genetic diversity (for A_R, H_E, H_O), whereas the site immediately downstream (HAY2) had the lowest genetic diversity (for H_E & H_O).

Population bottlenecks

With the BOTTLENECK program, the Wilcoxon one-tailed test for heterozygosity excess was significant for three of the Hayle sites (HAY1, HAY2, and HAY3), and also for other metal populations (RR1, RR2, and CRO) (Table 1). This suggests that these metal-impacted populations have experienced recent population declines. There was also evidence that the clean CAM2 and GAN2 populations have also bottlenecked, although these populations exhibited high overall genetic diversity.

M-ratio values suggested that bottlenecks had occurred in all populations—both metal-impacted and clean (Table 1). Across all sites and loci, the *M*-ratio ranged from 0.47 to 0.61, which in every case was lower than the calculated critical value (0.87). Furthermore, these values were also lower than the critical value of 0.68 proposed by Garza and Williamson (2001). This method did however seem to reflect the influence of metal contamination, as trout from metal-impacted sites had lower *M*-ratios compared to trout from clean sites (average 0.51 and 0.6, respectively).

Loci under selection

Tests for loci under selection using the Bayescan approach of Foll and Gaggiotti (2008), the *lnRV* statistic (Marshall and Weiss 2006), and the Fdist method (Excoffier et al. 2009) showed no reliably identifiable signals of diversifying selection in the various population group analyses (Fig. S2). Except for a signature of diversifying selection in the Red River populations for sasaTAP2A using the *lnRV* statistic, the MHC-linked loci showed little evidence of positive selection.

Genetic differentiation and population structure

Global F_{ST} for all populations and loci was 0.098. The highest F_{ST} between all populations was between HAY2 and CRO (F_{ST} = 0.106, P = 0.001, Table S3). The lowest F_{ST} was between the two downstream Hayle sites, HAY1 and HAY2 (F_{ST} = 0.006, P = 0.6). All pairwise F_{ST} values were statistically significant ($P < 0.05$), except for HAY1 and HAY2. Within the River Hayle, the global F_{ST} was 0.031. The highest F_{ST} was between the HAY2 and HAY4 sites, F_{ST} = 0.029 (P = 0.001), and the lowest was between HAY1 and HAY2, F_{ST} = 0.006 (P = 0.622). Across the Godolphin middle region, the pairwise F_{ST} between the downstream and upstream Hayle populations was 0.021 (P = 0.001).

Analysis of population structure using STRUCTURE showed that the most likely number of genetic groups was K = 3 (Fig. 2a; ΔK = 464.96): Group 1 (green) comprised CAM1, CAM2, GAN1, GAN2, FAL, TRES, RR1, and RR2; Group 2 (red) comprised the four Hayle populations (HAY1-HAY4); and Group 3 (blue) included the CRO, TREV1, and TREV2 populations (Fig. 2A). For Group 1, the hierarchical analyses showed ΔK = 3, whereby the groups were differentiated based on the river basin of origin, so that CAM1 & CAM2; GAN1 & GAN2; FAL & TRES, and RR1 & RR2 grouped together (Fig. 2B). Hierarchical analysis of Group 2 showed differentiation between the populations upstream and downstream of the highly contaminated Godolphin region of the Hayle (Fig. 2B). For Group 3, the hierarchical analysis separated the Crowlas from the two populations originating from the Trevaylor (Fig 2B).

The neighbor-joining dendrogram (Fig. 2C) revealed four population groups, each with high bootstrap support. The first group comprised the clean populations. Within this group, further structure was supported with high bootstraps, defining the population groups by river basin. The second group comprised the Hayle populations (HAY1-HAY4). The split between the upper and lower reaches of the Hayle was also evident. The third group includes the two sites from the Red River (RR1 & RR2), which cluster most closely to the Hayle but are still genetically distinct, despite their close geographical proximity. The final group forms the other three metal-impacted populations: the Crowlas (CRO), and the two sites on the Trevaylor (TREV1 & TREV2).

Dating genetic divergence

For the preliminary investigations (see Materials and Methods), all population-group comparisons showed that the posterior probability of Scenario 2 was higher, and therefore, it was more likely that population bottlenecks occurred prior to the generation of each metal-impacted population. This is further confirmed by the measures of genetic diversity (Table 1). Four scenarios were included in the final analysis (Fig. S1)—one scenario each from Groups 1 and 2 and two scenarios from Group 3 (as both scenarios had high logistic regression values). The most likely scenario was where metal-affected populations were derived

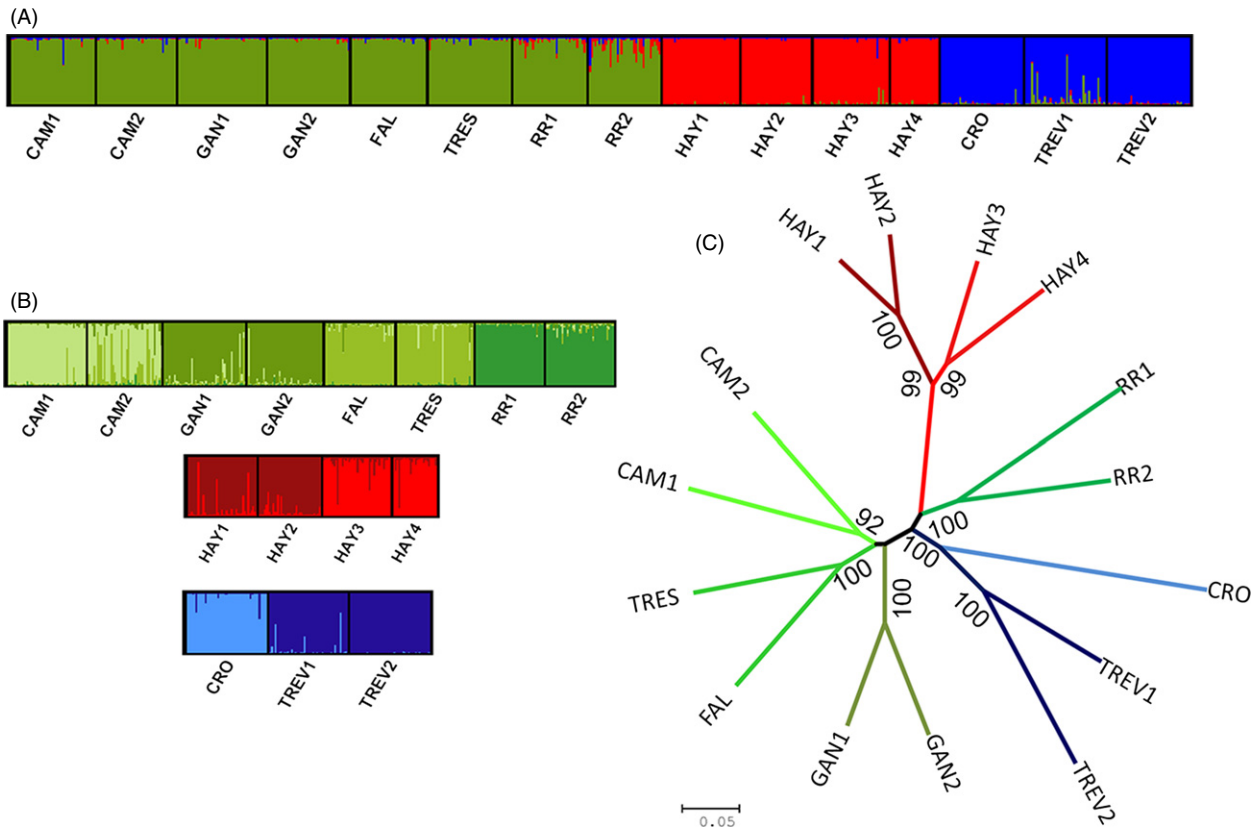


Figure 2 (A and B) Hierarchical STRUCTURE analyses showing estimated proportions of the coefficient of admixture of each individual's genome that originated from population *K*. (A) Primary STRUCTURE plot of all populations ($K = 3$). (B) From left to right; in green: hierarchical plot, where $K = 4$; in red: hierarchical plot, where $K = 2$; in blue: hierarchical plot, where $K = 3$. (C) Neighbor-joining phenogram, based on Cavalli-Sforza and Edwards chord distance (D_{CE}) showing the relationships between 15 populations of brown trout. Bootstrap values (% based on 1000 replicates) are show next to relevant branches; only bootstraps >90% are shown in the figure. Colors for each population match those presented in A and B.

multiple, independent times from a general clean lineage (logistic regression = 0.9721: Fig. S1: Group 2, Scenario 1). The mean mutation rate was 5.34×10^{-4} (95% CI: 3.23×10^{-4} to 8.71×10^{-4}). Using a generation time of 4 years, this split occurred approximately 960 years ago, c.1050 AD (t2 = 240 generations, 95% CI: 96.5–730; Table 2). After the split, the effective population size (N_E) of the Red River populations reduced by ~34% and the Trevaylor and Crowlas populations actually showed a small (~4%) increase in N_E. The subsequent split within the Hayle (t1) occurred approximately 155 years ago, c.1860 AD (t1 = 38.9 generations, 95% CI: 15.3–156). The River Hayle populations experienced a ~70% reduction in N_E. This reduction is the most severe of all population declines calculated (Table 2).

Table 2. Median values and 95% confidence intervals (CI) for DIYABC parameters for Scenario 1, Group 2 (See Fig. S1 for scenario topography).Values are in generations.

Parameter	Median	95% CI
N_{clean}	9220	7670–9930
$N_{Red\ river}$	3270	1430–6870
$N_{Crowlas,Trevaylor}$	5640	2960–8680
$N_{Hayle\ Downstream}$	1400	605–3590
$N_{Hayle\ Upstream}$	2900	1100–6900
$t1_{Hayle\ split}$	38.9	15.3–156
$t2_{Clean\text{-}metal\ split}$	240	96.5–730
$DB_{Red\ river}$	165	27.2–288
$N2_{Red\ river}$	4950	588–9480
$DB_{Crowlas,Trevaylor}$	159	24.9–286
$N2_{Crowlas,Trevaylor}$	5430	748–9570
DB_{Hayle}	180	29.6–290
$N2_{Hayle}$	4570	485–9430
μmic	5.34×10^{-4}	3.23×10^{-4} to 8.71×10^{-4}

N = effective population size after bottleneck, N2 = effective population size before bottleneck, DB = duration of bottleneck. μmic = mean mutation rate.

Discussion

The question of whether the genetic patterns of natural populations can be altered by human disturbance is a

salient issue in modern ecology and conservation. We used a panel of microsatellites to investigate the capacity of metal mining to act as a driver of genetic change in brown trout populations, on both a local and a regional scale. We explored both the genetic diversity and differentiation of trout occupying metal-impacted rivers compared to trout from relatively unpolluted control sites. Using approximate Bayesian computation (ABC), we have placed time points on the demographical changes of these populations to elucidate how present-day genetic patterns have been affected by historical anthropogenic interference.

The DIYABC analyses revealed several insights into the genetic makeup of contemporary trout populations in relation to increased historical mining activity. Firstly, our analyses revealed that the split between Hayle trout populations upstream and downstream of the highly contaminated Godolphin region of the river occurred approximately 155 years ago, c.1860. The peak exploitation of mines in the Godolphin region was from 1815 to 1850 (Atkinson 1994). Analysis of contemporary environmental data demonstrates a marked increase in copper and zinc at Godolphin (Fig. S3), and a study of the invertebrate communities within the River Hayle showed a marked decrease in species diversity in this region (Brown 1977). Thus, the present genetic structure appears indicative of rapid changes associated with significantly increased metal contaminants within the river during the Industrial Revolution. Moreover, the DIYABC results corroborate the results of population bottleneck analyses, demonstrating that Hayle trout have experienced severe population declines associated with this period of mining activity.

The second identified split (t2, Table 2) demonstrated that all trout populations from metal-contaminated rivers were derived from a single common ancestor approximately 960 years ago, c.1050—during the Medieval period from whence documentary evidence of tin mining in the region first dates (Lewis 1965). In particular, this coincides with an increased demand for metals in England, due to considerable population growth (Schofield and Vince 2003), and advancements in mining technology (Hatcher 1973). The mining method of this period was predominantly tin streaming (Gerrard 2000), a process that used considerable amounts of water (Lewis 1965). For example, John de Treeures complained of the tin miners of Cornwall, 'by reason of the great quantity of water they deluge the land where they work' (c. 1300s); complaints of this sort were numerous throughout the Medieval period, due to the wholesale trenching and excavating for alluvial deposits in the soil (Lewis 1965).

Tin streaming had a huge impact on environmental geochemistry (Pirrie et al. 2002). Due to the low solubility of tin, the net result would not have been a significant increase in the amount of tin in the system (Weber 1985), but a likely release of other metal contaminants by liberating them from the lode (Yim 1981). A further effect was that tin mining caused huge amounts of sediment to be remobilized and carried further downstream (Thorndycraft et al. 2004), which would have had further significant environmental impacts. Salmonids are particularly sensitive to suspended solids, as they affect gravel permeability and oxygen supply in embryos (Schindler Wildhaber et al. 2014), swimming performance and physiology (Berli et al. 2014), and predator avoidance (Louhi et al. 2011). The demonstration of population bottlenecks in the DIYABC suggests that this period had detrimental effects on the populations affected by these mining practices.

It is therefore likely that the mining activity of the Medieval period, through both increased metal contamination and the sediment effects of tin streaming, would have driven changes in trout population connectivity and in-river genetic structure. Later mining activity during the Industrial Revolution would have added additional population-level pressures on trout populations already shaped in part by the effects of Medieval mining practices resulting in the genetic patterns we observe here.

We found that metal-impacted populations have reduced genetic diversity compared to relatively unaffected control populations from clean rivers. Trout from the River Hayle and Red River have very low levels of genetic variation compared to populations with little metal contamination. An exception to this pattern is the higher level of genetic diversity observed in the Trevaylor 1 population. This may be attributed to other factors: a patchwork of metal contamination (Webb et al. 1978); a negative association between the level of genetic diversity and distance from the river mouth (Primmer et al. 2006); effects of asymmetric gene flow and metapopulation dynamics (Palstra et al. 2007); or genetic instability caused by temporal fluctuations of allele frequencies within the river (Jensen et al. 2005).

These low levels of genetic diversity are likely to be one of the most commonly observed effects of exposure to environmental contaminants (Bickham et al. 2000; Van Straalen and Timmermans 2002) with several studies demonstrating evidence of dramatic genetic reductions related to metal contamination (Bourret et al. 2008; Ungherese et al. 2010; Mussali-Galante et al. 2013). Such diversity-reducing events are likely the result of population bottlenecks, which involve a temporary reduction of population size and subsequently an increase in the rate of genetic drift. The program BOTTLENECK has been shown to more accurately detect recent signatures of declines in effective population size (Beebee and Rowe 2001), whereas the recovery time of the *M*-ratio suggests that this test is more indicative of historical reductions (Garza and

Williamson 2001) that persisted for a comparatively longer duration (Williamson-Natesan 2005).

Tests using BOTTLENECK showed that recent demographic declines have occurred in several of the metal-impacted populations. This is a clear genetic signature of how metal contamination may have caused local extinctions in the trout occupying these contaminated rivers. It should also be noted that two comparably clean-river populations also showed evidence of a bottleneck using this method, although the cause of these declines cannot be determined.

Low *M*-ratio values for all populations in this study, based on comparisons to the critical *M* value calculated here (0.87), an *M*-ratio of 0.76 calculated in other studies of salmonids (Dillane et al. 2008; Frazer and Russello 2013) and the critical *M* value of 0.68 suggested by Garza and Williamson (2001) suggest that all populations have suffered population declines. Geologically, much of the southwest of Britain is dominated by metal-bearing rocks (Dines 1956; Webb et al. 1978), and with a known mining history dating back to the Bronze Age, it is possible that historical mining activity may have caused population declines in these 'clean' rivers in the past. In giant Amazonian river turtles (*Podocnemis expansa),* tests for demographic declines using BOTTLENECK showed that about half of populations had declined, whereas *M*-ratio values suggested all populations had suffered long-term declines as a result of hunting pressure (Pearse et al. 2006).

Our population structure analyses showed strong evidence of genetic differentiation between clean populations and those affected by metal contamination. Both population structure analyses suggest that although the 'clean' populations are geographically distant, they cluster together and thus constitute a relatively homogeneous group. By comparison, the metal-impacted rivers are geographically proximate, yet the genetic differentiation between the various populations places them in distinct population groups.

To our knowledge, there are no physical barriers to fish movement within the rivers studied and tests for isolation by distance were nonsignificant (Fig. S4). Therefore, the generation of genetically distinct metal groups in this study is likely the consequence of population bottlenecks and extensive genetic drift within rivers and also, a possible disruption of gene flow between rivers due to local adaptation to metal contamination. Hayle river water has been shown to be toxic to metal-naïve fish (Durrant et al. 2011). Furthermore, considerable metal accumulation in Hayle trout tissues, as well as identification of upregulated pathways involved in metal detoxification and ion homeostasis, suggests adaptation may play a role (Uren-Webster et al. 2013). The genetic patterns may suggest that the complex of different metals and their varying contamination levels within the impacted rivers are driving different underlying genetic responses. Populations of yellow perch (*Perca flavescens*) from two major mining regions in Canada showed distinct population structuring and reduced genetic variation that was correlated with cadmium concentration, but no relationship was found with levels of copper contamination (Bourret et al. 2008).

We found no significant indication that any of the loci utilized here show consistent evidence for selection. However, the striking patterns of neutral genetic divergence may suggest genetic changes elsewhere in the genome may have occurred. For example, the ecotoxicological impact of the metals may have caused large areas of the genome ('genomic islands') to undergo selective sweeps (Nosil and Feder 2012), and the power of just a handful of genetic markers may not be sufficient to uncover this signal. In studies looking for footprints of selection in the freshwater adaptation of three-spine stickleback (*Gasterosteus aculeatus*), only 2.8% of loci were found to be under directional selection in Baltic populations (Mäkinen et al. 2008), while nine genomic regions with significantly differentiated F_{ST}, accounting for only ~2.5% of the entire genome, were found in Alaskan populations (Hohenlohe et al. 2010). An applied genomics approach might help to elucidate the genomewide effects of metal contamination.

The River Hayle provided an ideal system for identifying the effects of metal contamination on trout at a local scale. We revisit previous findings, which concluded that although Hayle water had a negative effect on metal-naïve fish, there were no reductions in genetic diversity and within-river variation was not structured on the basis of metal contamination (Durrant et al. 2011). Our genetic diversity calculations showed that Hayle trout have significantly reduced variation compared to all other populations studied (both metal-impacted and clean). Interestingly, genetic diversity was higher at the upstream sites, the opposite of what we might expect given that gene flow in salmonids tends to be downstream, therefore increasing genetic diversity in downstream sites (Griffiths et al. 2009). These unusual patterns of genetic diversity and structure may be due to the effects of Godolphin mine washout and any waterborne contaminants are likely to be moving with the flow of the river, enforcing higher ecotoxicological pressures on populations downstream of the contaminated region. Metal contamination is considerably higher downstream of Godolphin (total zinc ~570 μg/L; total copper 36 μg/L; total arsenic ~11 μg/L) compared to the upstream section of the catchment (total zinc ~36 μg/L; total copper ~6 μg/L; total arsenic ~3 μg/L).

Our analyses of the within-genetic structuring of the Hayle reveal that the upper and lower regions are highly spatially structured, despite being just 7.7 km apart. While olfaction is a critically important behavior in salmonids and is known to be negatively affected by metal ions

(Hansen et al. 1999; Baldwin et al. 2003), in this instance, the strength of the barrier presented by the highly metal-impacted Godolphin zone appears to act to maintain structure in spite of any possible reduction in fidelity (Tierney et al. 2010) of homing by returning adult trout caused by increased metal ion levels in the river water. Metal concentrations in the Godolphin section of the Hayle are extremely high (total zinc ~2512 μg/L; total copper ~471 μg/L; total arsenic ~99 μg/L) and are thus likely to be restricting gene flow and reinforcing in-river population structure, through a combination of avoidance behavior (Woodward et al. 1995) and direct mortality on fish attempting to traverse the Godolphin toxic zone. Thus, the impact of this chemical barrier appears to restrict fish movement to the same extent as physical barriers observed in some other systems (Hansen et al. 2014).

Conclusions

Evolutionary change can be associated with human activities, and such disturbance history has been suggested to be influential on the genetic architecture of natural populations (Stockwell et al. 2003; Banks et al. 2013). We have shown that mining for metals in the southwest region of Britain has resulted in striking patterns of genetic diversity and population structure both within and between river systems. These changes have arisen both during an early period of increased mining intensity—the Medieval period —and later during the Industrial Revolution. We demonstrate evidence that the observed neutral genetic divergence is due to population bottlenecks, disruptions of gene flow, and a likely increased rate of genetic drift. However, there is also a possibility that metal-impacted trout have developed a genetic adaptation—this cannot be explored here, but future studies should develop on understanding the broader genomic effects, as well as the physiology and molecular mechanisms of potential metal tolerance.

Acknowledgements

We would like to thank the Environment Agency, Westcountry Rivers Trust, and the University of Exeter for funding this project. We are very grateful to the members of the Environment Agency Southwest Region Fisheries teams who electrofished and collected trout fin clips. We would also like to thank Dr Bruce Stockley, Matt Healey, Giles Rickard, Janice Shears, and Philip Shears for support with electrofishing additional samples from the River Hayle and Red River. Thanks also to Dr Eduarda Santos, Dr Rod Wilson, Dr Duncan Pirrie, and Dr Nic Bury, for guidance on the direction and analyses of the paper and also to Dr Patrick Hamilton and Charles Ikediashi for assistance with laboratory work and analyses.

Data archiving statement

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.j3127

Microsatellite genotypes are available in Dryad (doi: 10.5061/dryad.j3127).

Data on metal contaminant levels are available through the Environment Agency South West Enquiries [SouthWestEnquiries@environment-agency.gov.uk].

Author contributions

J.R.S, R.A.K, and J.R.P designed the research; J.R.P and R.A.K carried out the laboratory work, genotyping, and analyses of the data. J.R.P and J.R.S wrote the paper.

Literature cited

Allendorf, F. W., O. Berry, and N. Ryman 2014. So long to genetic diversity, and thanks for all the fish. Molecular Ecology **23**:23–25.

Amos, W., and A. Balmford 2001. When does conservation genetics matter? Heredity **87**:257–265.

Atkinson, B. 1994. Mining Sites in Cornwall: V. 2. Dyllansow Truran, Redruth.

Austin, C. C., E. N. Rittmeyer, L. A. Oliver, J. O. Andermann, G. R. Zug, G. H. Rodda, and N. D. Jackson 2011. The bioinvasion of Guam: inferring geographic origin, pace, pattern and process of an invasive lizard (Carlia) in the Pacific using multi-locus genomic data. Biological Invasions **13**:1951–1967.

Baldwin, D. H., J. F. Sandahl, J. S. Labenia, and N. L. Scholz 2003. Sublethal effects of copper on coho salmon: impacts on nonoverlapping receptor pathways in the peripheral olfactory nervous system. Environmental Toxicology and Chemistry **22**:2266–2274.

Banks, S. C., G. J. Cary, A. L. Smith, I. D. Davies, D. A. Driscoll, A. M. Gill, D. B. Lindenmayer et al. 2013. How does ecological disturbance influence genetic diversity? Trends in Ecology & Evolution **28**:670–679.

Beaumont, M. A., and D. J. Balding 2004. Identifying adaptive genetic divergence among populations from genome scans. Molecular Ecology **13**:969–980.

Beebee, T., and G. Rowe 2001. Application of genetic bottleneck testing to the investigation of amphibian declines: a case study with Natterjack Toads. Conservation Biology **15**:266–270.

Berli, B. I., M. J. H. Gilbert, A. L. Ralph, K. B. Tierney, and P. Burkhardt-Holm 2014. Acute exposure to a common suspended sediment affects the swimming performance and physiology of juvenile salmonids. Comparative Biochemistry and Physiology Part A, Molecular & Integrative Physiology **176**:1–10.

Bickham, J. W., S. Sandhu, P. D. Hebert, L. Chikhi, and R. Athwal 2000. Effects of chemical contaminants on genetic diversity in natural populations: implications for biomonitoring and ecotoxicology. Mutation Research **463**:33–51.

Bourret, V., P. Couture, P. G. C. Campbell, and L. Bernatchez 2008. Evolutionary ecotoxicology of wild yellow perch (*Perca flavescens*) populations chronically exposed to a polymetallic gradient. Aquatic Toxicology **86**:76–90.

Brown, B. 1977. Effects of mine drainage on the River Hayle, Cornwall a) factors affecting concentrations of copper, zinc and iron in water, sediments and dominant invertebrate fauna. Hydrobiologia **52**:2–3.

Buckley, J. A. 2002. The Cornish Mining Industry: A Brief History. Tor Mark, Redruth.

Cairney, M., J. B. Taggart, and B. Hoyheim 2000. Characterization of microsatellite and minisatellite loci in Atlantic salmon (*Salmo salar* L.) and cross-species amplification in other salmonids. Molecular Ecology **9**:2175–2178.

Cavalli-Sforza, L. L., and A. W. Edwards 1967. Phylogenetic analysis. Models and estimation procedures. American Journal of Human Genetics **19**:233–257.

Clark, R. W., W. S. Brown, R. Stechert, and K. R. Zamudio 2010. Roads, interrupted dispersal, and genetic diversity in timber rattlesnakes. Conservation Biology **24**:1059–1069.

Cohen, S. 2002. Strong positive selection and habitat-specific amino acid substitution patterns in MHC from an Estuarine Fish Under Intense Pollution Stress. Molecular Biology and Evolution **19**:1870–1880.

Cornuet, J. M., and G. Luikart 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics **144**:2001–2014.

Cornuet, J. M., P. Pudlo, J. Veyssier, A. Dehne-Garcia, M. Gautier, R. Leblois, J. M. Marin et al. 2014. DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism. DNA sequence and microsatellite data. Bioinformatics **30**:1187–1189.

Cornuet, J. M., F. Santos, M. A. Beaumont, C. P. Robert, J. M. Marin, D. J. Balding, T. Guillemaud et al. 2008. Inferring population history with DIYABC: a user-friendly approach to approximate Bayesian computation. Bioinformatics **24**:2713–2719.

Dillane, E., P. McGinnity, J. P. Coughlan, M. C. Cross, E. De Eyto, E. Kenchington, P. Prodöhl et al. 2008. Demographics and landscape features determine intrariver population structure in Atlantic salmon (*Salmo salar* L.): the case of the River Moy in Ireland. Molecular Ecology **17**:4786–4800.

Dines, H. G. 1956. The Metalliferous Mining Region of South-West England, Vol. **2**. H.M. Stationery Office, London.

Durrant, C. J., J. R. Stevens, C. Hogstrand, and N. R. Bury 2011. The effect of metal pollution on the population genetic structure of brown trout (*Salmo trutta* L.) residing in the River Hayle, Cornwall, UK. Environmental Pollution **159**:3595–3603.

Evanno, G., S. Regnaut, and J. Goudet 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology **14**:2611–2620.

Excoffier, L., T. Hofer, and M. Foll 2009. Detecting loci under selection in a hierarchically structured population. Heredity **103**:285–298.

Excoffier, L., and H. E. L. Lischer 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources **10**:564–567.

Foll, M., and O. Gaggiotti 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics **180**:977–993.

Fontaine, M. C., A. Snirc, A. Frantzis, E. Koutrakis, B. Oztürk, A. A. Oztürk, and F. Austerlitz 2012. History of expansion and anthropogenic collapse in a top marine predator of the Black Sea estimated from genetic data. Proceedings of the National Academy of Sciences of the United States of America **109**:E2569–E2576.

Frazer, K. K., and M. A. Russello 2013. Lack of parallel genetic patterns underlying the repeated ecological divergence of beach and stream-spawning kokanee salmon. Journal of Evolutionary Biology **26**:2606–2621.

Garza, J. C., and E. G. Williamson 2001. Detection of reduction in population size using data from microsatellite loci. Molecular Ecology **10**:305–318.

Gerrard, S. 2000. The Early British tin Industry. Tempus Publishing Ltd, Stroud.

Gharbi, K., A. Gautier, R. G. Danzmann, S. Gharbi, T. Sakamoto, B. Høyheim, J. B. Taggart et al. 2006. A linkage map for brown trout (*Salmo trutta*): chromosome homeologies and comparative genome organization with other salmonid fish. Genetics **172**:2405–2419.

González-Porter, G. P., F. Hailer, O. Flores-Villela, R. García-Anleu, and J. E. Maldonado 2011. Patterns of genetic diversity in the critically endangered Central American river turtle: human influence since the Mayan age? Conservation Genetics **12**:1229–1242.

Goudet, J. 1995. FSTAT (version 1.2): a computer program to calculate F-statistics. Journal of Heredity **86**:485–486.

Griffiths, A. M., I. Koizumi, D. Bright, and J. R. Stevens 2009. A case of isolation by distance and short-term temporal stability of population structure in brown trout (*Salmo trutta*) within the River Dart, South-west England. Evolutionary Applications **2**:537–554.

Grimholt, U., F. Drabløs, S. M. Jørgensen, B. Høyheim, and R. J. M. Stet 2002. The major histocompatibility class I locus in Atlantic salmon (*Salmo salar* L.): polymorphism, linkage analysis and protein modelling. Immunogenetics **54**:570–581.

Guillemaud, T., M. A. Beaumont, M. Ciosi, J.-M. Corneut, and A. Estoup 2010. Inferring introduction routes of invasive species using approximate Bayesian computation on microsatellite data. Heredity **104**:88–99.

Han, F. X., A. Banin, Y. Su, D. L. Monts, M. J. Plodinec, W. L. Kingery, and G. E. Triplett 2002. Industrial age anthropogenic inputs of heavy metals into the pedosphere. Die Naturwissenschaften **89**:497–504.

Hansen, M. M., M. T. Limborg, A.-L. Ferchaud, and J.-M. Pujolar 2014. The effects of Medieval dams on genetic divergence and demographic history in brown trout populations. BMC Evolutionary Biology **14**:122.

Hansen, B. H., S. Rømma, Ø. A. Garmo, P. A. Olsvik, and R. A. Andersen 2006a. Antioxidative stress proteins and their gene expression in brown trout (*Salmo trutta*) from three rivers with different heavy metal levels. Comparative Biochemistry and Physiology Toxicology & Pharmacology **143**:263–274.

Hansen, B. H., S. Rømma, L. I. R. Søfteland, P. A. Olsvik, and R. A. Andersen 2006b. Induction and activity of oxidative stress-related proteins during waterborne Cu-exposure in brown trout (*Salmo trutta*). Chemosphere **65**:1707–1714.

Hansen, J. A., J. D. Rose, R. A. Jenkins, K. G. Gerow, and H. L. Bergman 1999. Chinook salmon (*Oncorhynchus tshawytscha*) and rainbow trout (*Oncorhynchus mykiss*) exposed to copper: Neurophysiological and histological effects on the olfactory system. Environmental Toxicology and Chemistry **18**:1979–1991.

Hatcher, J. 1973. English tin Production and Trade Before 1550. Clarendon Press, Oxford.

Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson, and W. A. Cresko 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genetics **6**: e1000862.

Jensen, L. F., M. M. Hansen, J. Carlsson, V. Loeschcke, and K.-L. D. Mensberg 2005. Spatial and temporal genetic differentiation and effective population size of brown trout (*Salmo trutta,* L.) in small Danish rivers. Conservation Genetics **6**:615–621.

Jensen, L. F., M. M. Hansen, C. Pertoldi, G. Holdensgaard, K. L. Mensberg, and V. Loeschcke 2008. Local adaptation in brown trout early

life-history traits: implications for climate change adaptability. Proceedings of the Royal Society B: Biological Sciences **275**:2859–2868.

Jones, O. R., and J. Wang 2010. COLONY: a program for parentage and sibship inference from multilocus genotype data. Molecular Ecology Resources **10**:551–555.

Keenan, K., C. R. Bradley, J. J. Magee, R. A. Hynes, R. J. Kennedy, W. W. Crozier, R. Poole et al. 2013. Beaufort trout MicroPlex: a high-throughput multiplex platform comprising 38 informative microsatellite loci for use in resident and anadromous (sea trout) brown trout *Salmo trutta* genetic studies. Journal of Fish Biology **82**:1789–1804.

King, T. L., M. S. Eackles, and B. H. Letcher 2005. Microsatellite DNA markers for the study of Atlantic salmon (*Salmo salar*) kinship, population structure, and mixed-fishery analyses. Molecular Ecology Notes **5**:130–132.

Kyle, C. J., Y. Rico, S. Castillo, V. Srithayakumar, C. I. Cullingham, B. N. White, and B. A. Pond 2014. Spatial patterns of neutral and functional genetic variations reveal patterns of local adaptation in raccoon (*Procyon lotor*) populations exposed to raccoon rabies. Molecular Ecology **23**:2287–2298.

Lewis, G. R. 1965. The Stanneries: A Study of Medieval Tin Miners of Cornwall. D. Bradford Barton Ltd., Truro.

Lind, E. E., and M. Grahn 2011. Directional genetic selection by pulp mill effluent on multiple natural populations of three-spined stickleback (*Gasterosteus aculeatus*). Ecotoxicology **20**:503–512.

Louhi, P., M. Ovaska, A. Mäki-Petäys, J. Erkinaro, and T. Muotka 2011. Does fine sediment constrain salmonid alevin development and survival? Canadian Journal of Fisheries and Aquatic Sciences **68**:1819–1826.

Lunn, D. J., A. Thomas, N. Best, and D. Spiegelhalter 2000. WinBUGS – A Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing **10**:325–337.

Mäkinen, H. S., J. M. Cano, and J. Merilä 2008. Identifying footprints of directional and balancing selection in marine and freshwater three-spined stickleback (*Gasterosteus aculeatus*) populations. Molecular Ecology **17**:3565–3582.

Marshall, J. M., and R. E. Weiss 2006. A Bayesian heterogeneous analysis of variance approach to inferring recent selective sweeps. Genetics **173**:2357–2370.

Martinez, J. L., P. Moran, and E. Garcia-Vazquez 1999. Dinucleotide repeat polymorphism at the SS4, SS6 and SS11 loci in atlantic salmon (*Salmo salar*). Animal Genetics **30**:462–478.

McFarlane, D. A., J. Lundberg, and H. Neff 2013. A speleothem record of early British and roman mining at charterhouse, Mendip, England. Archaeometry **56**:431–443.

Mondol, S., M. W. Bruford, and U. Ramakrishnan 2013. Demographic loss, genetic structure and the conservation implications for Indian tigers. Proceedings of the Royal Society B: Biological Sciences **280**:20130496.

Mussali-Galante, P., E. Tovar-Sánchez, M. Valverde, L. Valencia-Cuevas, and E. Rojas 2013. Evidence of population genetic effects in *Peromyscus melanophrys* chronically exposed to mine tailings in Morelos, Mexico. Environmental Science and Pollution Research International **20**:7666–7679.

Nosil, P., and J. L. Feder 2012. Genomic divergence during speciation: causes and consequences. Philosophical Transactions of the Royal Society of London Series B, Biological Sciences **367**:332–342.

O'Reilly, P. T., L. C. Hamilton, S. K. McConnell, and J. M. Wright 1996. Rapid analysis of genetic variation in Atlantic salmon (*Salmo salar*) by PCR multiplexing of dinucleotide and tetranucleotide microsatellites. Canadian Journal of Fisheries and Aquatic Sciences **53**:2292–2298.

Olsen, J. B., S. L. Wilson, E. J. Kretschmer, K. C. Jones, and J. E. Seeb 2000. Characterization of 14 tetranucleotide microsatellite loci derived from sockeye salmon. Molecular Ecology **9**:2185–2187.

van Oosterhout, C., W. F. Hutchinson, D. P. M. Wills, and P. Shipley 2004. Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes **4**:535–538.

Palstra, F. P., M. F. O'Connell, and D. E. Ruzzante 2007. Population structure and gene flow reversals in Atlantic salmon (*Salmo salar*) over contemporary and long-term temporal scales: effects of population size and life history. Molecular Ecology **16**:4504–4522.

Paterson, S., S. B. Piertney, D. Knox, J. Gilbey, and E. Verspoor 2004. Characterization and PCR multiplexing of novel highly variable tetranucleotide Atlantic salmon (*Salmo salar* L.) microsatellites. Molecular Ecology Notes **4**:160–162.

Peakall, R., and P. E. Smouse 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research–an update. Bioinformatics **28**:2537–2539.

Pearse, D. E., A. D. Arndt, N. Valenzuela, B. A. Miller, V. Cantarelli, and J. W. Jr Sites 2006. Estimating population structure under nonequilibrium conditions in a conservation context: continent-wide population genetics of the giant Amazon river turtle, *Podocnemis expansa* (Chelonia; Podocnemididae). Molecular Ecology **15**:985–1006.

Peery, M. Z., R. Kirby, B. N. Reid, R. Stoelting, E. Doucet-Bëer, S. Robinson, C. Vásquez-Carrillo et al. 2012. Reliability of genetic bottleneck tests for detecting recent population declines. Molecular Ecology **21**:3403–3418.

Penhallurick, R. D. 1986. Tin in Antiquity: Its Mining and Trade Throughout the Ancient World With Particular Reference to Cornwall / Trove. Institute of Materials, Minerals and Mining, London.

Pirrie, D., M. R. Power, P. D. Wheeler, A. B. Cundy, C. Bridges, and G. Davey 2002. Geochemical signature of historical mining: Fowey Estuary, Cornwall, UK. Journal of Geochemical Exploration **76**:31–43.

Primmer, C. R., A. J. Veselov, A. Zubchenko, A. Poututkin, I. Bakhmet, and M. T. Koskinen 2006. Isolation by distance within a river system: genetic population structuring of Atlantic salmon, *Salmo salar*, in tributaries of the Varzuga River in northwest Russia. Molecular Ecology **15**:653–666.

Pritchard, J. K., M. Stephens, and P. Donnelly 2000. Inference of population structure using multilocus genotype data. Genetics **155**:945–959.

Rainbow, P. S., S. Kriefman, B. D. Smith, and S. N. Luoma 2011. Have the bioavailabilities of trace metals to a suite of biomonitors changed over three decades in SW England estuaries historically affected by mining? The Science of the Total Environment **409**:1589–1602.

Rousset, F. 2008. Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Molecular Ecology Resources **8**:103–106.

Sánchez, J. A., C. Clabby, D. Ramos, G. Blanco, F. Flavin, E. Vázquez, and R. Powell 1996. Protein and microsatellite single locus variability in *Salmo salar* L. (Atlantic salmon). Heredity **77**:423–432.

Schindler Wildhaber, Y., C. Michel, J. Epting, R. A. Wildhaber, E. Huber, P. Huggenberger, P. Burkhardt-Holm et al. 2014. Effects of river morphology, hydraulic gradients, and sediment deposition on water exchange and oxygen dynamics in salmonid redds. The Science of the Total Environment **470–471**:488–500.

Schlotterer, C. 2002. A microsatellite-based multilocus screen for the identification of local selective sweeps. Genetics **160**:753–763.

Schlötterer, C., and D. Dieringer 2005. A novel test statistic for the identification of local selective sweeps based on microsatellite gene diversity. In D. Nurminsky, ed. Selective Sweep, pp. 55–64. Molecular Biology Intelligence Unit, New York.

Schofield, J., and A. G. Vince 2003. Medieval Towns: The Archaeology of British Towns in Their European Setting. A&C Black, London.

Slettan, A., I. Olsaker, and Ø. Lie 1995. Atlantic salmon, *Salmo salar*, microsatellites at the SSOSL25, SSOSL85, SSOSL311, SSOSL417 loci. Animal Genetics **26**:281–282.

Smith, T., and L. Bernatchez 2008. Evolutionary change in human-altered environments. Molecular Ecology **17**:1–8.

Stockwell, C. A., A. P. Hendry, and M. T. Kinnison 2003. Contemporary evolution meets conservation biology. Trends in Ecology & Evolution **18**:94–101.

Storey, J. D., and R. Tibshirani 2003. Statistical significance for genome-wide studies. Proceedings of the National Academy of Sciences of the United States of America **100**:9440–9445.

Van Straalen, N., and M. J. Timmermans 2002. Genetic variation in toxicant-stressed populations: an evaluation of the "genetic erosion" hypothesis. Human and Ecological Risk Assessment **8**:983–1002.

Tamura, K., G. Stecher, D. Peterson, A. Filipski, and S. Kumar 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Molecular Biology and Evolution **30**:2725–2729.

Thorndycraft, V. R., D. Pirrie, and A. G. Brown 2004. Alluvial records of medieval and prehistoric tin mining on Dartmoor, southwest England. Geoarchaeology **19**:219–236.

Tierney, K. B., D. H. Baldwin, T. J. Hara, P. S. Ross, N. L. Scholz, and C. J. Kennedy 2010. Olfactory toxicity in fishes. Aquatic Toxicology **96**:2–26.

Truett, G. E., P. Heeger, R. L. Mynatt, A. A. Truett, J. A. Walker, and M. L. Warman 2000. Preparation of PCR-quality mouse genomic DNA with hot sodium hydroxide and tris (HotSHOT). BioTechniques **29**:52–54.

Ungherese, G., A. Mengoni, S. Somigli, D. Baroni, S. Focardi, and A. Ugolini 2010. Relationship between heavy metals pollution and genetic diversity in Mediterranean populations of the sandhopper *Talitrus saltator* (Montagu) (Crustacea, Amphipoda). Environmental Pollution **158**:1638–1643.

Uren Webster, T. M., N. Bury, R. van Aerle, and E. M. Santos 2013. Global transcriptome profiling reveals molecular mechanisms of metal tolerance in a chronically exposed wild population of brown trout. Environmental Science and Technology **47**:8869–8877.

Vähä, J.-P., J. Erkinaro, E. Niemelä, and C. R. Primmer 2007. Life-history and habitat features influence the within-river genetic structure of Atlantic salmon. Molecular Ecology **16**:2638–2654.

Vasemägi, A., J. Nilsson, and C. R. Primmer 2005. Seventy-five EST-linked Atlantic salmon (*Salmo salar* L.) microsatellite markers and their cross-amplification in five salmonid species. Molecular Ecology Notes **5**:282–288.

Webb, J., I. Thornton, M. Thompson, R. Howarth, and P. Lowenstein 1978. The Wolfson Geochemical Atlas of England and Wales. Clarendon Press, Oxford.

Weber, G. 1985. The importance of tin in the environment and its determination at trace levels. Fresenius' Zeitschrift fur Analytische Chemie **321**:217–224.

Wedepohl, K. 1991. The composition of the upper earth's crust and the natural cycles of selected metals; Metals in natural raw materials. Natural resources. In: E. Merian, ed. Metals and Their Compounds in the Environment: Occurrence, Analysis and Biological Relevance, pp. 3–17. VCH, Weinheim.

Weeks, A. R., C. M. Sgro, A. G. Young, R. Frankham, N. J. Mitchell, K. A. Miller, M. Byrne et al. 2011. Assessing the benefits and risks of translocations in changing environments: a genetic perspective. Evolutionary Applications **4**:709–725.

Williamson-Natesan, E. G. 2005. Comparison of methods for detecting bottlenecks from microsatellite loci. Conservation Genetics **6**:551–562.

Wood, C. 2011. An introduction to metals in fish physiology and toxicology: Basic Principles. In C. Wood, A. Farrell, and C. Braune, eds. Fish Physiology: Homeostasis and Toxicology of Essential Metals: Homeostasis and Toxicology of Essential Metals, pp. 2–40. Academic Press – Elsevier, London, Waltham, San Diego.

Woodward, D. F., J. A. Hansen, H. L. Bergman, A. J. DeLonay, and E. E. Little 1995. Brown trout avoidance of metals in water characteristic of the Clark Fork River, Montana. Canadian Journal of Fisheries and Aquatic Sciences **52**:2031–2037.

Yim, W. W.-S. 1981. Geochemical investigations on fluvial sediments contaminated by tin-mine tailings, Cornwall, England. Environmental Geology **3**:245–256.

Zelikoff, J. T. 1993. Metal pollution-induced immunomodulation in fish. Toxicological Sciences **22**:1–7.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. DIYABC scenarios tested on the various population groups.

Figure S2. Tests for loci under selection using Bayescan (A), *lnRV* (B) and Fdist (C).

Figure S3. from the Environment Agency (1990–2014) on concentrations of metals (μg/L) at the sites across the River Hayle.

Figure S4. Relationship between genetic distance (FST/(1-FST)) and geographic distance (kilometres) for the 15 populations of brown trout, based on 99 permutations; rxy = −0.09, $P = 0.28$, $R^2 = 0.0082$.

Table S1. Location and site identification for each sampled population.

Table S2. Details of the eight multiplexes including quantities of forward and reverse primer, total volumes and PCR touchdown (TD) programme.

Table S3. Pairwise F_{ST} results for each of the 15 brown trout populations.

Supporting Information

Human mining activity across the ages determines the genetic structure of modern brown trout (*Salmo trutta*) populations

This supporting information contains:

Table S1. Location and site identification for each sampled population.

Table S2. Details of the eight multiplexes including quantities of forward and reverse primer, total volumes and PCR touchdown (TD) programme.

Table S3. Pairwise F_{ST} results for each of the 15 brown trout populations.

Figure S1. DIYABC scenarios tested on the various population groups.

Figure S2. Tests for loci under selection using Bayescan (A), *lnRV* (B) and Fdist (C).

Figure S3. Data from the Environment Agency (1990–2014) on concentrations of metals (µg/L) at the sites across the River Hayle.

Figure S4. Relationship between genetic distance ($F_{ST}/(1-F_{ST})$) and geographic distance (kilometres) for the 15 populations of brown trout, based on 99 permutations; rxy = −0.09, $P = 0.28$, $R^2 = 0.0082$.

Supporting Information: Table 1. Location and site identification for each sampled population. N - number of individuals sampled. Coordinates in decimal degrees (WGS84). The populations are grouped into 'clean' sites, where there is little contemporary metal-contamination, and 'metal-impacted sites', which are currently negatively affected by metal pollution. Site codes match those presented in Figure 1.

	Catchment	Sub-catchment	Site	Site code	N	Coordinates
Control sites	Camel	Allen	Trehannick	CAM1	49	50.579, -4.733
	Camel	Stannon Stream	Stannon	CAM2	44	50.594, -4.687
	Gannel	Main River	Gwills	GAN1	50	50.393, -5.058
	Gannel	Main River	Kestle Mill	GAN2	50	50.393, -5.024
	Fal	Main River	Tregony	FAL	47	50.267, -4.918
	Tresillian	Main River	Geen Mill	TRES	48	50.286, -4.981
Metal-impacted sites	Red River	Tehidy Brook	Tehidy Brook	RR1	45	50.233, -5.325
	Red River	Main River	Roseworthy farm	RR2	41	50.20, -5.341
	Hayle	Main River	St Erth	HAY1	44	50.164, -5.433
	Hayle	Main River	Porthcollum	HAY2	48	50.149, -5.421
	Hayle	Main River	Drym farm	HAY3	48	50.149, -5.335
	Hayle	Main River	Clowance wood	HAY4	37	50.157, -5.330
	Coastal Streams	Crowlas	Cuccurian	CRO	49	50.159, -5.494
	Trevaylor	Main River	Trythogga	TREV1	50	50.127, -5.526
	Trevaylor	Main River	Noongallas	TREV2	50	50.145, -5.551

Supporting Information. Table 2: Details of the eight multiplexes including quantities of forward and reverse primer, total volumes and PCR touchdown (TD) programme. For each numerically identified multiplex, groups a and b were pooled together for analysis on the Beckman Coulter sequencer. *MHC class 1 linked loci

Multiplex	Locus	Reference	F	R	Volumes	PCR Program
MP1a	Ssa85	O'Reilly *et al.* 1996	1μM	1μM	28μM Primer	TD: 60°-50°
	Ssa52NVH	Gharbi *et al.* 2006	8μM	8μM	72μM H20	
	SsosL417	Slettan *et al.* 1995	1μM	1μM		
	SS11	Martinez *et al.* 1999	4μM	4μM		
MP1b	BG935488	Vasemägi *et al.* 2005	1.5μM	1.5μM	33μM Primer	TD: 60°-50°
	CA048828	Vasemägi *et al.* 2005	6μM	6μM	67μM H20	
	CA060177	Vasemägi *et al.* 2005	8μM	8μM		
	CA060208	Vasemägi *et al.* 2005	1μM	1μM		
MP2a	SsaD157	King *et al.* 2005	4μM	4μM	30μM Primer	TD: 55°-45°
	SsaD58	King *et al.* 2005	6μM	6μM	70μM H20	
	SsaF43	Sánchez *et al.* 1996	1μM	1μM		
	SsosL311	Slettan *et al.* 1995	4μM	4μM		
MP2b	Str3QUB	Keenan *et al.* 2013	1μM	1μM	14μM Primer	TD: 60°-50°
	Ssa407UOS	Cairney *et al.* 2000	6μM	6μM	86μM H20	
MP3a	Ssa197	O'Reilly *et al.*1996	1μM	1μM	14μM Primer	TD: 60°-50°
	sasaTAP2A*	Grimholt *et al.* 2002	2μM	2μM		
MP3b	SSsp2213	Paterson *et al.* 2004	4μM	4μM	14μM Primer	TD: 55°-45°
	One102	Olsen *et al.* 2000	3μM	3μM	86μM H20	
MP4a	CA053293	Vasemägi *et al.* 2005	8μM	8μM	60μM Primer	TD: 60°-50°
	sasa-UBA*	Grimholt *et al.* 2002	6μM	6μM	40μM H20	
	CA769358	Vasemägi *et al.* 2005	4μM	4μM		
	CA515794	Vasemägi *et al.* 2005	8μM	8μM		
	sasaTAP2B*	Grimholt *et al.* 2002	4μM	4μM		
MP4b	Ssa412UOS	Cairney *et al.* 2000	5μM	5μM	10μM Primer	TD: 60°-50°

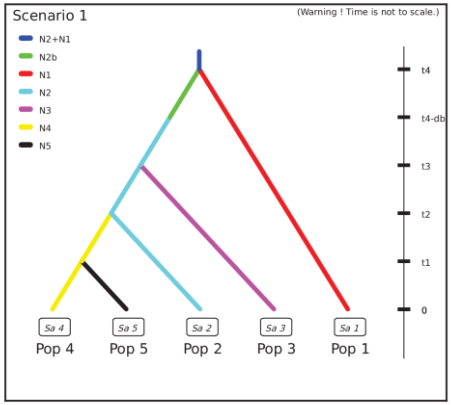
Supporting Information: Table S3. Pairwise F_{ST} results for each of the 15 brown trout populations. F_{ST} values lie below the black diagonal boxes, significance values are above the black diagonal boxes: * = p<0.05, ** = p<0.01, *** = p<0.001.

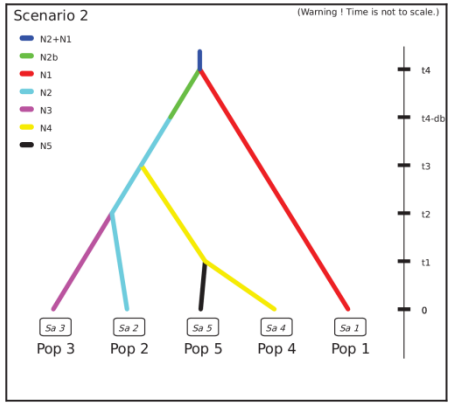
CAM2	GAN1	GAN2	FAL	TRES	RR1	RR2	HAY1	HAY2	HAY3	HAY4	CRO	TREV1	TREV2	
***	***	***	***	***	***	***	***	***	***	***	***	***	***	CAM1
0.000	***	***	***	***	***	***	***	***	***	***	***	***	***	CAM2
0.022	0.000	*	***	***	***	***	***	***	***	***	***	***	***	GAN1
0.023	0.007	0.000	***	***	***	***	***	***	***	***	***	***	***	GAN2
0.019	0.027	0.031	0.000	***	***	***	***	***	***	***	***	***	***	FAL
0.020	0.024	0.028	0.014	0.000	***	***	***	***	***	***	***	***	***	TRES
0.045	0.046	0.047	0.050	0.045	0.000	***	***	***	***	***	***	***	***	RR1
0.039	0.040	0.038	0.043	0.045	0.030	0.000	***	***	***	***	***	***	***	RR2
0.071	0.073	0.071	0.077	0.081	0.093	0.073	0.000	0.602	***	***	***	***	***	HAY1
0.073	0.073	0.073	0.081	0.084	0.093	0.075	0.006	0.000	***	***	***	***	***	HAY2
0.057	0.055	0.057	0.060	0.060	0.071	0.058	0.024	0.026	0.000	**	***	***	***	HAY3
0.063	0.061	0.062	0.069	0.066	0.076	0.063	0.027	0.029	0.012	0.000	***	***	***	HAY4
0.060	0.071	0.066	0.063	0.063	0.073	0.060	0.103	0.106	0.084	0.094	0.000	***	***	CRO
0.036	0.040	0.039	0.038	0.037	0.055	0.047	0.082	0.086	0.062	0.069	0.048	0.000	***	TREV1
0.059	0.058	0.058	0.060	0.056	0.069	0.068	0.099	0.101	0.077	0.080	0.063	0.022	0.000	TREV2

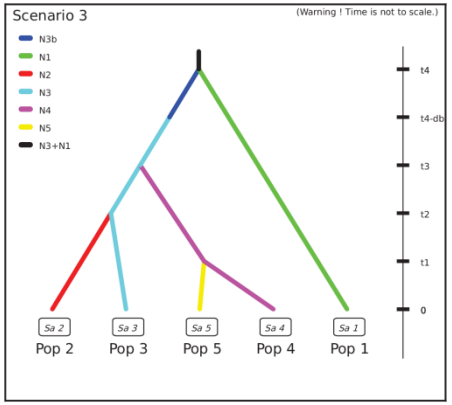
Supporting Information: Figure 1. DIYABC scenarios tested on the various population groups. Group 1: populations are derived multiple times from a common ancestor. Group 2: metal populations are derived multiple times from a general clean lineage. Group 3: the clean and metal groups are separate lineages, with populations from metal contaminated sites being derived from the common metal-contaminated lineage. Scenarios surrounded by a red box were shown to be the most likely, based on logistic regression of the posterior probabilities of 10^5 simulations per scenario. These four scenarios were compared against one another in further simulations (10^6 simulations per scenario).

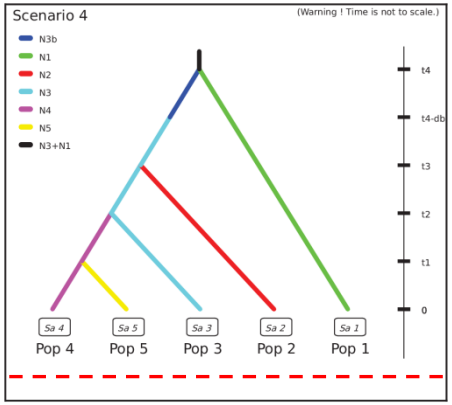
Group 1

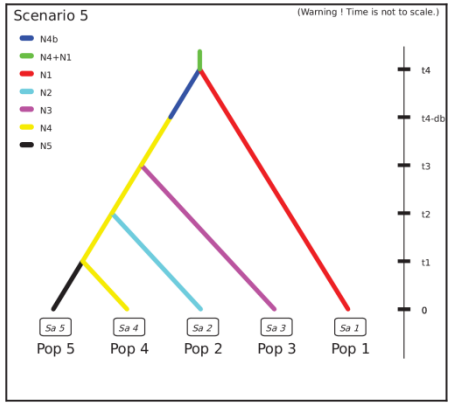
Common ancestor scenarios

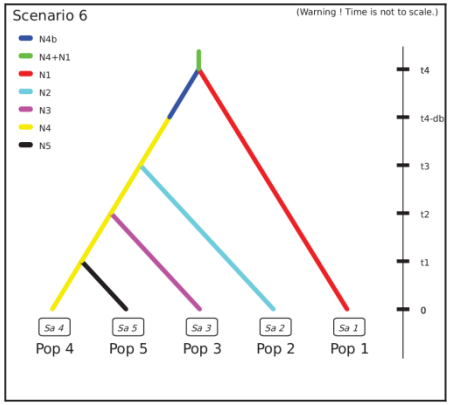


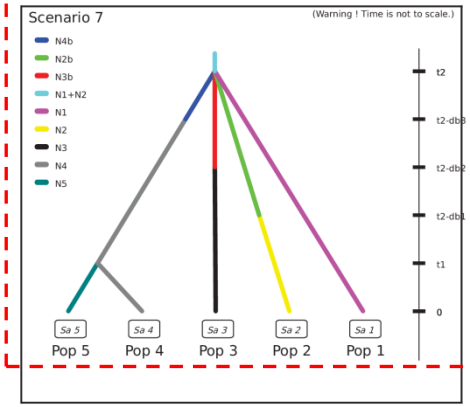






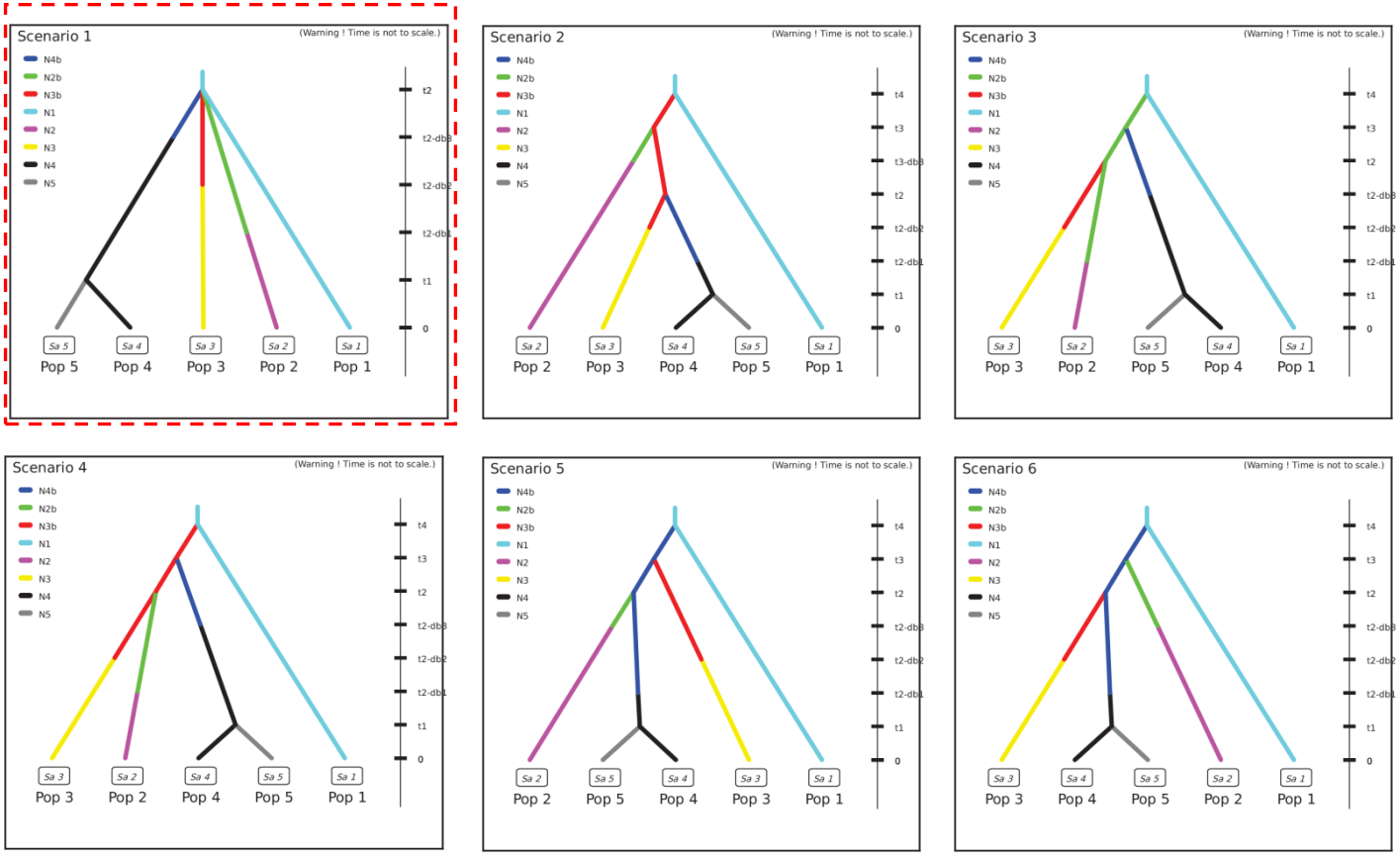






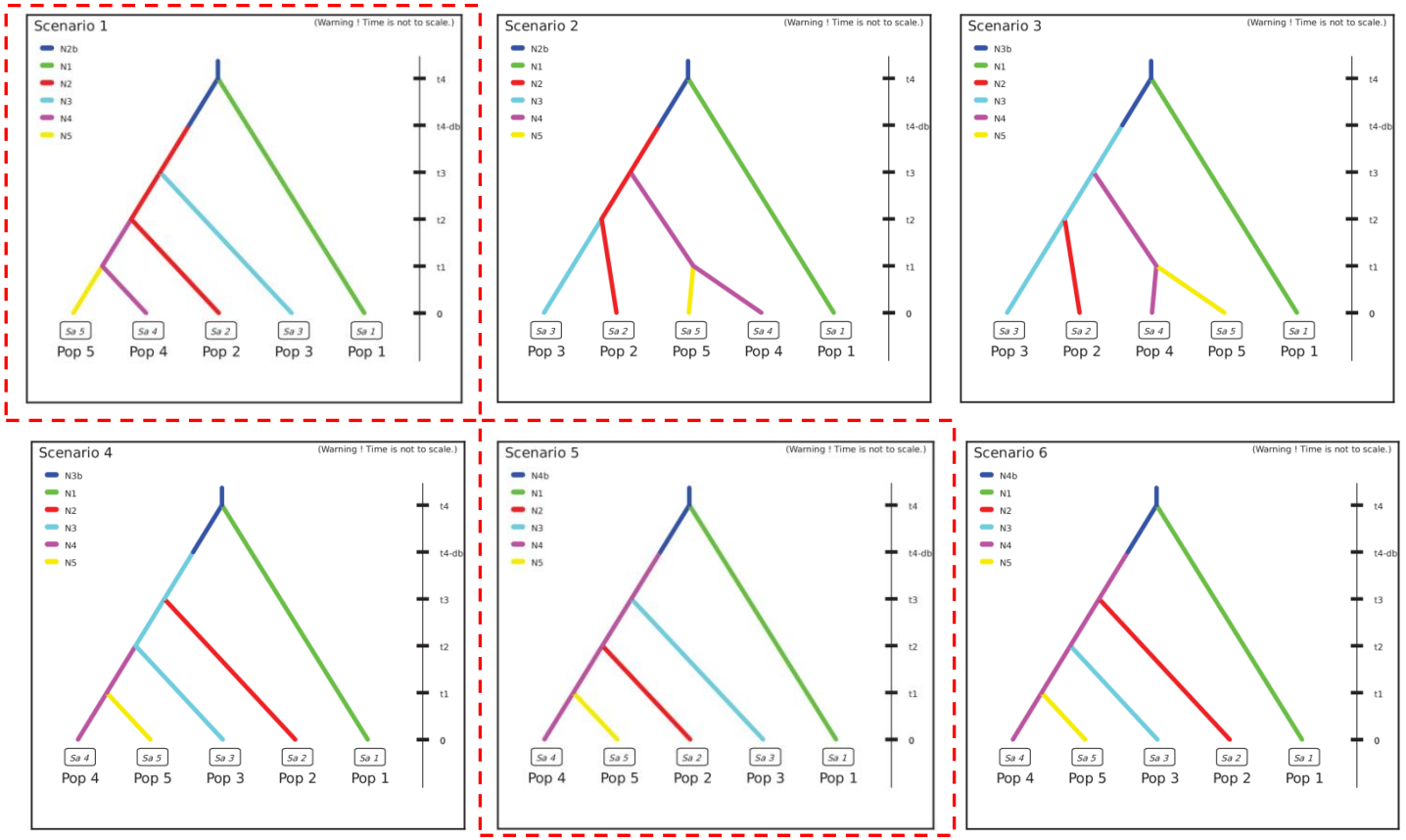
Group 2

Clean lineage scenarios

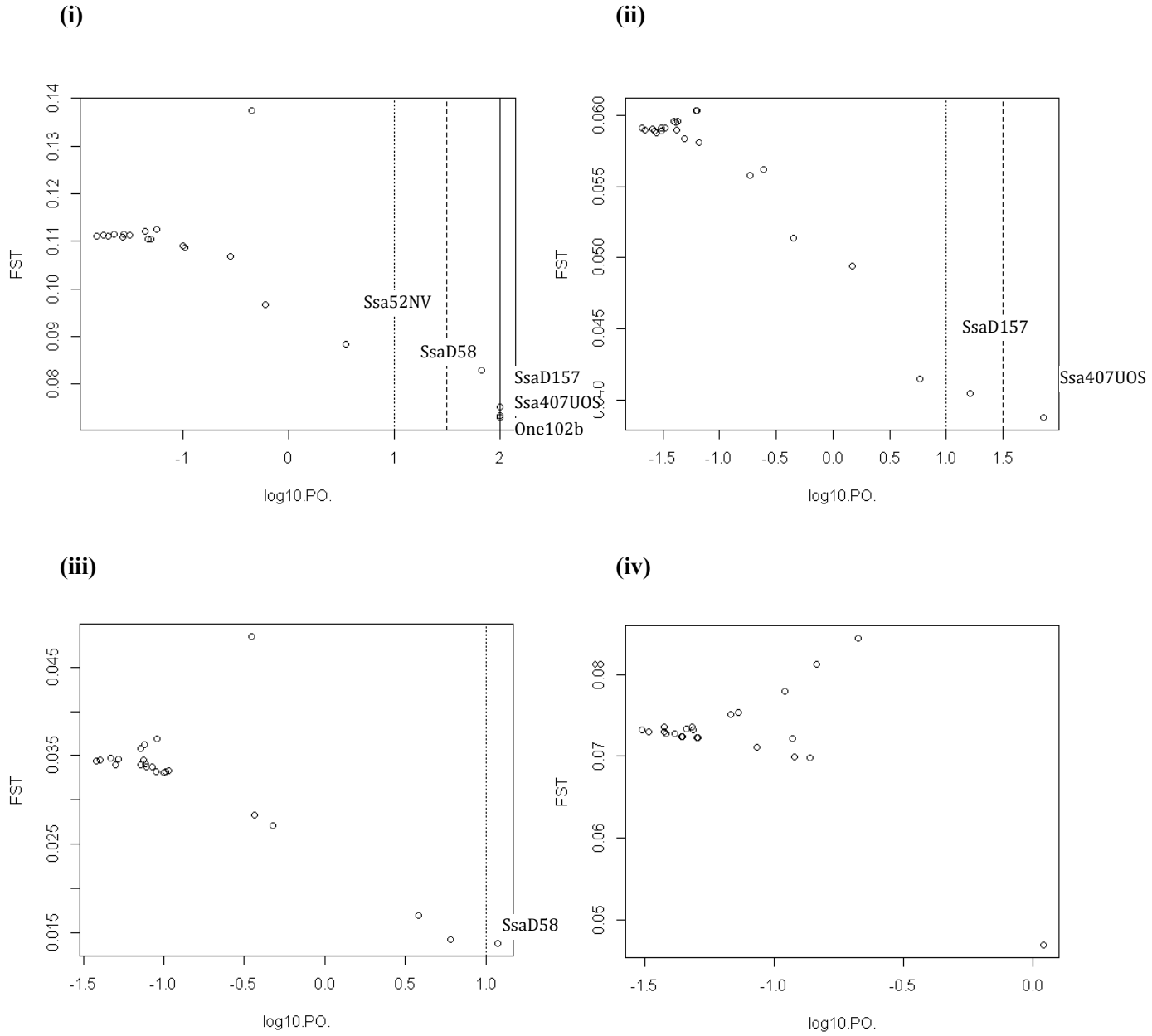


Group 3

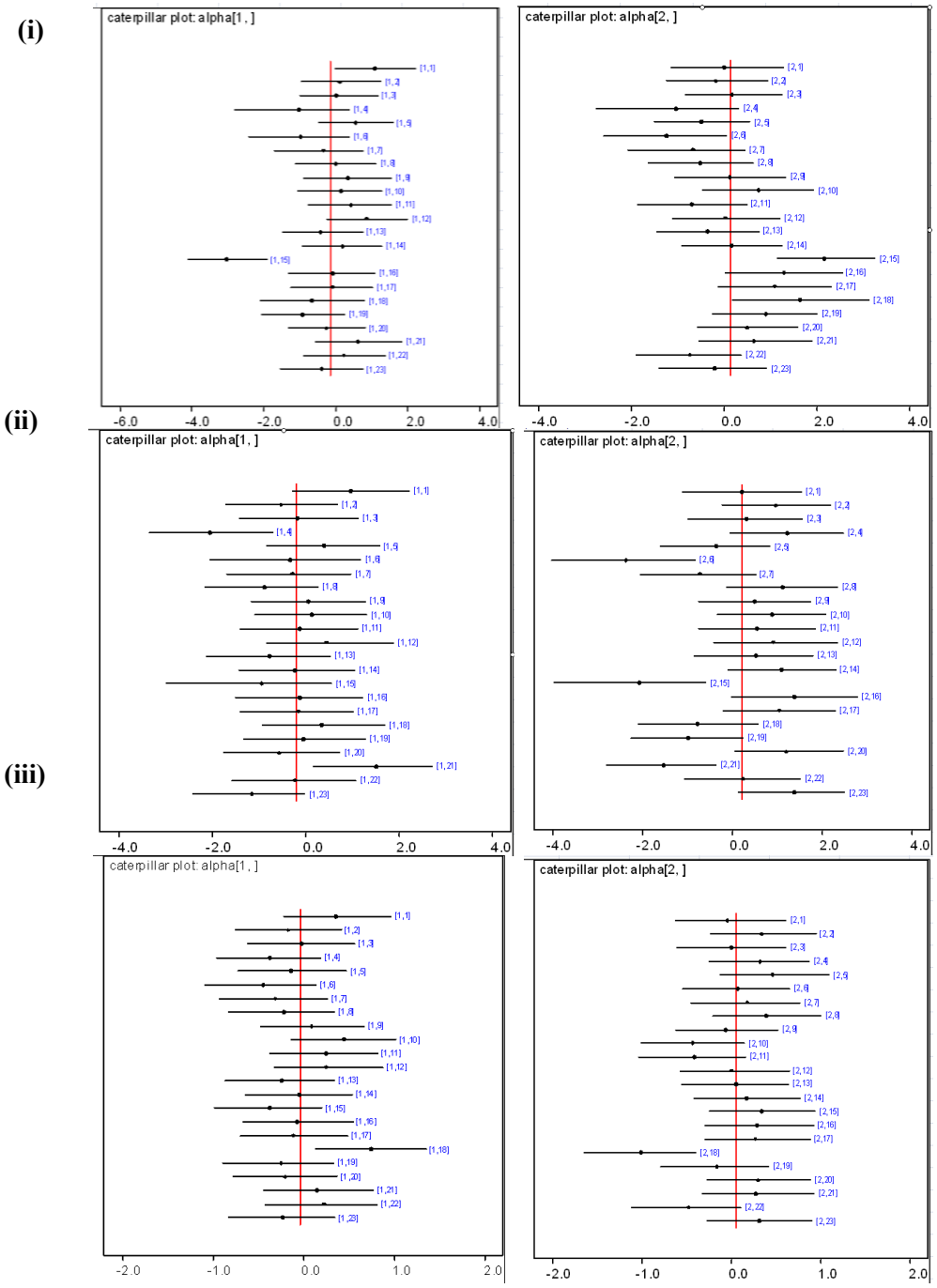
Metal lineage scenarios



Supporting Information: Figure 2. Tests for loci under selection using Bayescan (**2A**), *lnRV* (**2B**) and Fdist (**2C**)



Supplementary Information: Figure 2A. Bayescan results using FDIST method. Log10 Posterior Odds (PO) value against F_{ST}. Vertical lines represent cut-off values for posterior odds: dotted line >1; dashed line >1.5; solid line > 2. Some loci analysed have a Log10 Posterior Odds value of 1000; these are represented on the graph as 2. (i) All populations. (ii) Clean populations. (iii) Hayle populations. (iv) Crowlas and Trevaylor populations.



Supplementary Information: Figure 2B. Caterpillar plots showing posterior distributions of α_{ij}-values used to detect selective sweeps using the *lnRV* statistic. Median values of $\alpha_{ij} > 0.5$ are indicative of selection. We considered α_{ij}-values >1 to be indicative of strong selection. (i) Comparison 1 – Clean vs. Hayle populations. Loci with high α_{ij}-values: One102a & Ssa85. (ii) Comparison 2 – Clean vs. CROW/TREV1/TREV2. Loci with high α_{ij}-values: Ssa85, CA060208, One102a, Ssa412 & TAP2B. (iii) Comparison 3 – Clean vs. Red River populations. Loci with high α_{ij}-values: sasaTAP2A. Caterpillar plot bounds represent lower (2.5%) and upper (97.5%) quartiles of the Bayesian interval.

Supplementary Information: Figure 2C. Table of results from the Fdist test, implemented in Arlequin,. (i) Comparison 1 - Clean vs. Hayle populations. (ii) Comparison 2 – Clean vs. CROW/TREV1/TREV2. (iii) Comparison 3 – Clean vs. Red River populations. (iv) Comparison 4 – Clean vs. all metal populations. p values <0.01 (in red) are significant.

(i)

Locus	Ho	Fct	P-value run 1	P-value run 2	P-value run 3
Ssa52NVH	1.0031	0.1328	0.0094	0.0117	0.0106
CA048828	0.9631	0.0727	0.1638	0.1700	0.1720
BG935488	0.7620	0.0314	0.1823	0.1826	0.1835
CA060208	0.6338	0.0661	0.4513	0.4570	0.4351
SS11	0.8578	0.1249	0.2056	0.1949	0.2011
Ssa85	0.6412	0.1615	0.1437	0.1488	0.1399
SsosL417	0.8863	0.0363	0.1736	0.1832	0.1715
CA060177	0.8627	0.0344	0.1572	0.1614	0.1491
SsosL311	0.9498	0.0589	0.3400	0.3483	0.3485
Ssa407UOS	0.9708	0.0557	0.2310	0.2324	0.2412
Str3QUB	0.5164	0.0398	0.3402	0.3355	0.3252
SsaD58	0.9743	0.0525	0.2385	0.2383	0.2476
SsaF43	0.8127	0.1280	0.2099	0.2273	0.2199
SsaD157	0.9776	0.0666	0.1380	0.1403	0.1450
One102a	0.4853	0.2349	0.0400	0.0430	0.0492
One102b	0.9325	0.0321	0.2086	0.2153	0.2063
SSsp2213	0.9223	0.0749	0.3756	0.3686	0.3687
sasaTAP2A	0.8471	0.1297	0.1937	0.1897	0.1930
Ssa197	0.8906	0.1763	0.0614	0.0556	0.0577
CA515794	0.7988	0.0584	0.3837	0.3712	0.3781
Ssa412UOS	0.7939	0.2163	0.0513	0.0656	0.0593
CA769358	0.8262	0.0551	0.3551	0.3414	0.3328
sasaTAP2B	0.6014	0.0827	0.3738	0.3747	0.3666

(ii)

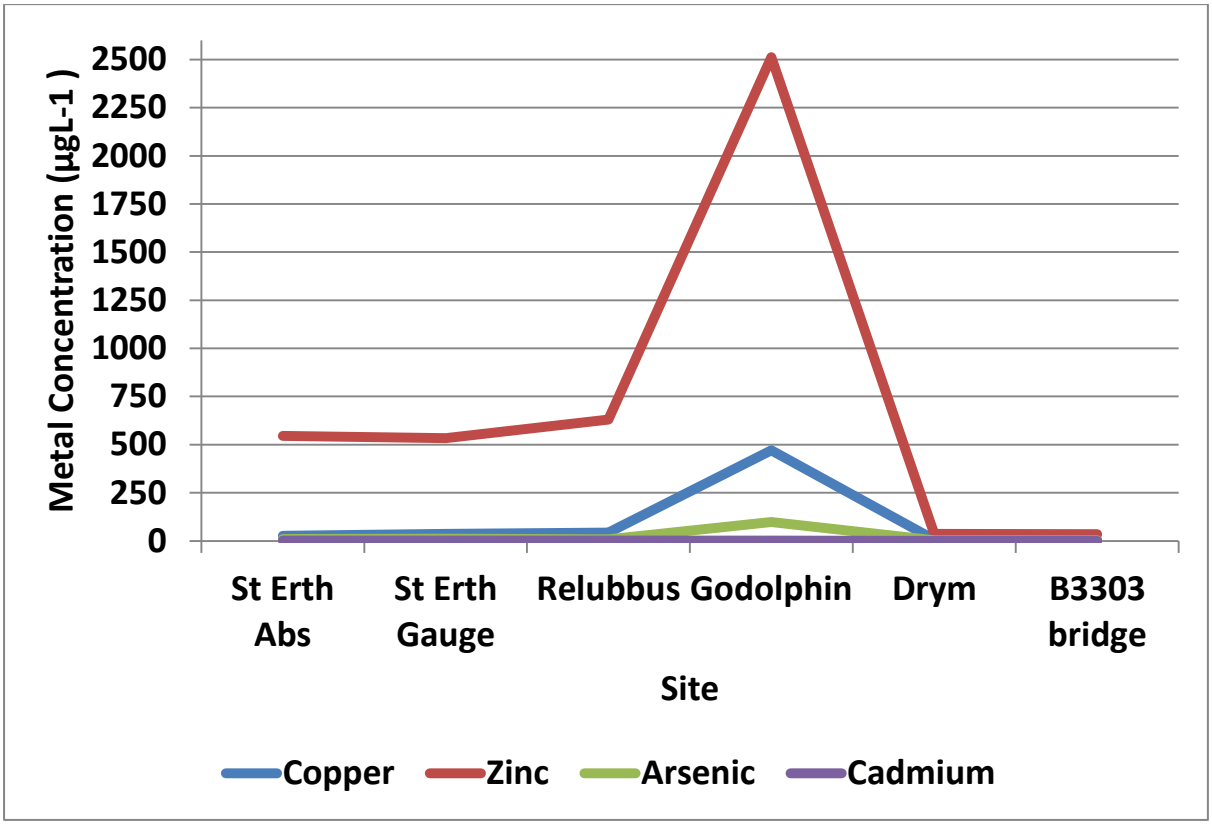
Locus	Ho	FCT	P-value run 1	P-value run 2	P-value run 3
Ssa52NVH	0.9677	0.0219	0.4227	0.4189	0.4221
CA048828	0.9558	0.0583	0.0821	0.0838	0.0860
BG935488	0.8212	0.0332	0.3946	0.3921	0.4017
CA060208	0.7407	0.0934	0.1205	0.1150	0.1229
SS11	0.9116	0.0497	0.2893	0.2823	0.2981
Ssa85	0.5555	0.1960	0.0325	0.0395	0.0356
SsosL417	0.8896	0.0249	0.2861	0.2877	0.2825
CA060177	0.8893	0.0484	0.3343	0.3249	0.3387
SsosL311	0.9305	0.0009	0.0063	0.0060	0.0059
Ssa407UOS	0.9627	0.0188	0.3397	0.3362	0.3391
Str3QUB	0.5622	0.0101	0.1217	0.1191	0.1209
SsaD58	0.9840	0.0559	0.0366	0.0375	0.0393
SsaF43	0.8121	0.0432	0.4122	0.4142	0.3949
SsaD157	0.9603	0.0223	0.4075	0.4028	0.4052
One102a	0.4878	0.1822	0.0258	0.0256	0.0231
One102b	0.9331	0.0327	0.4675	0.4720	0.4645
SSsp2213	0.9163	0.0336	0.4518	0.4530	0.4400
sasaTAP2A	0.7903	-0.0094	0.0364	0.0349	0.0381
Ssa197	0.7899	0.0187	0.1992	0.2067	0.2127
CA515794	0.8685	0.0448	0.3842	0.3836	0.3862
Ssa412UOS	0.7993	0.0133	0.1337	0.1362	0.1427
CA769358	0.7995	0.0955	0.1087	0.1059	0.1019
sasaTAP2B	0.5551	0.0195	0.2250	0.2176	0.2219

(iii)

Locus	HO	FCT	P-value run 1	P-value run 2	P-value run 3
Ssa52NVH	0.9800	0.0399	0.1249	0.1180	0.1193
CA048828	0.9583	0.0872	0.0115	0.0131	0.0126
BG935488	0.7728	0.0242	0.3114	0.3057	0.3060
CA060208	0.6112	0.0148	0.1927	0.1839	0.1868
SS11	0.8398	0.0254	0.3297	0.3372	0.3297
Ssa85	0.5742	0.0536	0.2647	0.2789	0.2735
SsosL417	0.9027	0.0496	0.2375	0.2483	0.2409
CA060177	0.8776	0.0401	0.3647	0.3767	0.3693
SsosL311	0.9756	0.0680	0.0212	0.0231	0.0230
Ssa407UOS	0.9743	0.0307	0.2562	0.2443	0.2483
Str3QUB	0.6356	0.0527	0.2531	0.2463	0.2612
SsaD58	0.9734	0.0376	0.1642	0.1553	0.1573
SsaF43	0.7882	0.0076	0.0848	0.0839	0.0836
SsaD157	0.9747	0.0515	0.0640	0.0622	0.0616
One102a	0.4876	0.0337	0.3003	0.3010	0.2914
One102b	0.9217	0.0336	0.4198	0.4162	0.4181
SSsp2213	0.9247	0.0595	0.1273	0.1294	0.1265
sasaTAP2A	0.7880	-0.0094	0.0544	0.0512	0.0547
Ssa197	0.7878	0.0050	0.0543	0.0537	0.0535
CA515794	0.8055	0.0019	0.0187	0.0191	0.0186
Ssa412UOS	0.8217	0.0592	0.2224	0.2048	0.2137
CA769358	0.8343	0.0444	0.3341	0.3213	0.3339
sasaTAP2B	0.5645	0.0184	0.2248	0.2154	0.2226

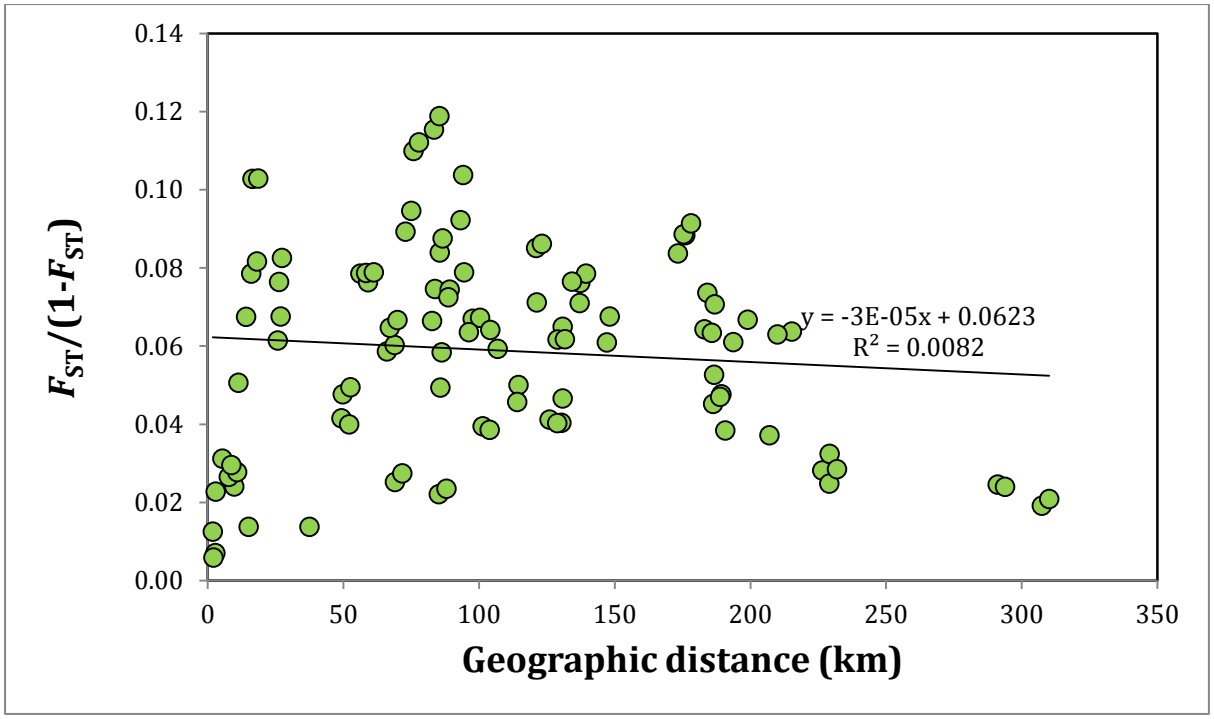
(iv)

Locus	Ho	FCT	P-value run 1	P-value run 2	P-value run 3
Ssa52NVH	0.9602	0.0471	0.0393	0.0433	0.0407
CA048828	0.9437	0.0475	0.0704	0.0721	0.0710
BG935488	0.7810	0.0025	0.0339	0.0330	0.0340
CA060208	0.6803	0.0398	0.2672	0.2613	0.2708
SS11	0.8525	0.0133	0.2022	0.2011	0.2057
Ssa85	0.5532	-0.0338	0.0015	0.0017	0.0019
SsosL417	0.8682	0.0318	0.3491	0.3453	0.3426
CA060177	0.8729	0.0209	0.3462	0.3484	0.3510
SsosL311	0.9343	0.0189	0.3579	0.3582	0.3538
Ssa407UOS	0.9604	0.0116	0.2498	0.2464	0.2465
Str3QUB	0.5392	-0.0018	0.2384	0.2442	0.2395
SsaD58	0.9649	0.0245	0.2712	0.2787	0.2769
SsaF43	0.8014	0.0487	0.1808	0.1911	0.1878
SsaD157	0.9612	0.0314	0.1608	0.1679	0.1660
One102a	0.4243	0.1921	0.0068	0.0106	0.0067
One102b	0.9192	0.0042	0.0553	0.0556	0.0542
SSsp2213	0.9055	0.0377	0.2400	0.2353	0.2443
sasaTAP2A	0.8009	0.0244	0.3911	0.3829	0.3900
Ssa197	0.8193	0.0308	0.3658	0.3741	0.3715
CA515794	0.8235	0.0116	0.1735	0.1700	0.1726
Ssa412UOS	0.7374	0.0757	0.0773	0.0807	0.0785
CA769358	0.7847	0.0106	0.1612	0.1562	0.1614
sasaTAP2B	0.5739	-0.0103	0.1207	0.1206	0.1120

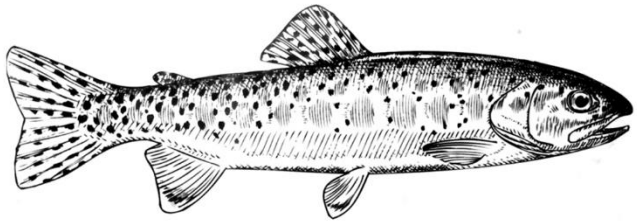


Site	Copper	Zinc	Arsenic	Cadmium
St Erth Abs	27.38	545.75	10.92	1.13
St Erth Gauge	36.74	533.85	12.27	1.19
Relubbus	42.68	630.8	8.82	1.38
Godolphin	471.15	2511.59	98.68	2.58
Drym	6.34	35.99	2.97	0.46
B3303 bridge	4.97	35.95	3.66	0.36

Supporting Information: Figure 3. Data from the Environment Agency (1990-2014) on concentrations of metals (µgL -1) at the sites across the River Hayle. In relation to the sites listed in Table 1 - HAY1: St Erth Abstraction (Abs) and St Erth Gauge. HAY2: Relubbus. Godolphin middle region. HAY3: Drym. HAY4: B3303 bridge.



Supplementary Information: Figure 4. Relationship between genetic distance ($F_{ST}/(1-F_{ST})$) and geographic distance (kilometres) for the 15 populations of brown trout, based on 99 permutations; rxy=-0.09, P=0.28, R^2=0.0082.



Chapter III

Lost in parameter space: A road map for Stacks

Published in *Methods in Ecology and Evolution* (2017). DOI: 10.1111/2041-210X.12775

AUTHORS: J.R. Paris[1], J.R. Stevens[1] and J.M. Catchen[2]

[1] Biosciences, College of Life and Environmental Sciences, University of Exeter, Exeter, UK

[2] Department of Animal Biology, University of Illinois at Urbana–Champaign, Urbana, IL 61801

J.R. Paris conceived and implemented the analytical methods, and wrote the manuscript

J.R. Stevens conceived of the brown trout experiments and improved the manuscript

J.M. Catchen conceived and implemented the analytical methods, and improved the manuscript

Methods in Ecology and Evolution

BRITISH ECOLOGICAL SOCIETY

Methods in Ecology and Evolution 2017 doi: 10.1111/2041-210X.12775

Lost in parameter space: a road map for STACKS

Josephine R. Paris[1], Jamie R. Stevens[1] and Julian M. Catchen*[,2]

[1]*Biosciences, College of Life and Environmental Sciences, University of Exeter, Exeter, UK; and* [2]*Department of Animal Biology, University of Illinois at Urbana–Champaign, Urbana, IL 61801, USA*

Summary

1. Restriction site-Associated DNA sequencing (RAD-seq) has become a widely adopted method for genotyping populations of model and non-model organisms. Generating a reliable set of loci for downstream analysis requires appropriate use of bioinformatics software, such as the program STACKS.

2. Using three empirical RAD-seq datasets, we demonstrate a method for optimising a *de novo* assembly of loci using STACKS. By iterating values of the program's main parameters and plotting resultant core metrics for visualisation, researchers can gain a much better understanding of their dataset and select an optimal set of parameters; we present the *80% rule* as a generally effective method to select the core parameters for STACKS.

3. Visualisation of the metrics plotted for the three RAD-seq datasets shows that they differ in the optimal parameters that should be used to maximise the amount of available biological information. We also demonstrate that building loci *de novo* and then integrating alignment positions is more effective than aligning raw reads directly to a reference genome.

4. Our methods will help the community in honing the analytical skills necessary to accurately assemble a RAD-seq dataset.

Key-words: alignment, *de novo* assembly, parameter optimisation, population genetics, RAD-seq, STACKS

Introduction

The last decade has been punctuated by innovations in the generation of genomic data for evolutionary and ecological science. The development of massively parallel, short-read sequencing, associated with lowered costs and open-access analysis tools, has enabled the genomic interrogation of a multitude of model and non-model species. Restriction site-Associated DNA sequencing (RAD-seq) has been proven to be an effective method for identifying and screening high-resolution polymorphism within and between populations (Lescak *et al.* 2015; Blanco-Bercial & Bucklin 2016), ecotypes (Hale *et al.* 2013; Pavey *et al.* 2015) and species (Nadeau *et al.* 2012; Wagner *et al.* 2013; Pante *et al.* 2015).

Restriction site-Associated DNA sequencing involves creating a reduced representation of a genome by isolating the DNA connected to a set of restriction enzyme cut sites. This cost-effective approach can be repeated in large numbers of samples to produce nearly the same reduced subset of the genome in each individual. After sequencing, the data are reassembled into loci, anchored by the presence of the restriction enzyme cut site (Baird *et al.* 2008; Etter *et al.* 2011), and subsequently SNPs are identified across those loci. Many different flavours of the original RAD protocol are now used (e.g. ddRAD, ezRAD, GBS, 2bRAD) and several analysis programs exist (reviewed in Andrews *et al.* 2016). Restriction site-Associated DNA sequencing provides a highly flexible experimental approach, which can be tuned by choosing restriction enzymes with different properties, such as cutting frequency, or by choosing combinations of enzymes; however, this flexibility also brings challenges, such as the quality of DNA required and ascertainment bias stemming from natural variation in restriction enzyme cut sites across a set of populations or species. A number of studies have outlined these aspects of experimental design and technical considerations (Davey *et al.* 2011, 2013; Rowe, Renaut & Guggisberg 2011; Arnold *et al.* 2013).

Restriction site-Associated DNA sequencing is facilitating a shift from using scores of genetic markers to make biological inferences (e.g. microsatellites), to using large-scale data obtained from tens of thousands of loci. The transition from familiar and well-established genotyping techniques into more complex genomic analyses remains a daunting task for many researchers. Thus, the ability to correctly handle the bioinformatic analysis of these vastly larger datasets is essential. Error quantification and hierarchical methods for filtering datasets to obtain biologically robust RAD-seq data do exist. However, these techniques require additional sequencing effort (Mastretta-Yanes *et al.* 2015) and particular RAD datasets for which testing the error is feasible (Fountain *et al.* 2016). Moreover, a RAD-seq analysis relies on competent bioinformatics knowledge, all of which are demanding, especially in a cost- or

*Correspondence author. E-mail: jcatchen@illinois.edu

Table 1. (Top) Three main parameters that control locus formation and polymorphism in STACKS, the default values, the STACKS component program that uses the parameter and a description of what part of the processes each parameter controls. (Bottom) Four additional parameters referenced in the paper (but not part of the optimisation process)

Parameter	Default value	STACKS component	Description
m	3	`ustacks`	Minimum number of raw reads required to form a stack (a putative allele)
M	2	`ustacks`	Number of mismatches allowed between stacks (putative alleles) to merge them into a putative locus
n	1	`cstacks`	Number of mismatches allowed between stacks (putative loci) during construction of the catalog
N	*M* + 2	`ustacks`	Number of mismatches allowed to align secondary reads (reads that did not form stacks) to assembled putative loci to increase locus depth
r	0	`populations`	Percentage of individuals that must possess a particular locus for it to be included in calculation of population-level statistics
max_obs_het	1	`populations`	For a particular locus the maximum number of heterozygous individuals that may be present
min_maf	0	`populations`	For a particular locus, alleles occurring below this frequency are discarded

time-limited framework. To this end, a fundamental protocol for developing an accurate and economical RAD-seq data interrogation strategy is currently lacking.

One of the most widely used programs for processing RAD-seq data is STACKS (Catchen *et al*. 2011, 2013a), a software pipeline designed to assemble loci from short-read sequences derived from restriction enzyme-based protocols, such as RAD-seq. STACKS can be used to assemble markers for genetic mapping analyses (Amores *et al*. 2011), or for population genomics (Epstein *et al*. 2016; Laporte *et al*. 2016), phylogeography (Emerson *et al*. 2010; Bryson *et al*. 2016), and phylogenomics (Jones *et al*. 2013; Díaz-Arce *et al*. 2016). The popularity of STACKS lies in its versatility and user-driven application. After cleaning and demultiplexing raw data (`process_radtags`), the researcher can proceed through one of two main pipelines depending on the availability of a reference genome, to build loci either *de novo* (`denovo_map.pl`; hereafter *de novo map*) or reference aligned (`ref_map.pl`; hereafter *ref map*). Throughout an analysis, values must be chosen for key parameters, which frequently have a significant effect on the building and quality of the resulting loci.

Stack assembly is controlled by several main parameters, the choice of which will depend on key features of a RAD-seq dataset: (i) Biological, such as the inherent polymorphism, level of ploidy and the biological hypothesis being tested; (ii) Study dependent, such as the number of individuals multiplexed, RAD flavour (e.g. RAD, ddRAD, 2bRAD) and restriction enzyme used, including the number of cut sites, the number and length of raw reads, coverage, sequencing platform and inherent error and (iii) Library development issues, for example, degraded DNA (Graham *et al*. 2015) and exogenous contamination (Trucchi *et al*. 2016). Given the uniqueness of each dataset, choosing which parameters are optimal for stack assembly can be difficult.

We tested and optimised the three main parameters within the *de novo map* pipeline (Table 1), which determine the number and the polymorphism of loci in a RAD-seq dataset. The first two affect how loci are built within each individual sample using the core component `ustacks`, where *m* is the minimum number of raw reads required to form a stack (or putative allele) and *M* is the number of mismatches allowed between stacks to merge them into a putative locus (Catchen *et al*. 2011, 2013a).

After the building of loci at an individual level, `cstacks` attempts to match loci across samples to build a *catalog*, which represents the homologous loci across all population samples. To accommodate fixed differences in loci between individuals, mismatches are also allowed during the construction of the *catalog* and the number allowed is controlled by the *n* parameter. Here, we outline a method where we demonstrate that by iterating a range of values for the main parameters, followed by plotting core assembly metrics gathered from the STACKS output files, the researcher can observe and make an informed choice of the best parameter sets for their data (Table 2).

An attractive attribute of RAD-seq is that it can be adopted in model and non-model organisms alike, by using either a reference genome or by constructing loci *de novo*. Both approaches require optimisation of the parameter space; in the case of a *de novo* assembly the parameters must be supplied to STACKS directly, whereas in a reference-aligned analysis the analogous parameters must be supplied to the chosen aligner. An advantage of a *de novo* assembly is that STACKS will identify putative alleles one after another and then merge them into putative loci – leveraging biological information – while an aligner will independently align each raw read. The reference genome also acts as a filter; for example, a draft genome will exclude loci not contained in the assembly and may fail to align reads that belong to loci captured in the reference more than once (e.g. as haplotypes). Alternatively, a reference genome of the study species may not be available, but the genome of a closely related species may be, providing positional information for loci at a reduced precision. We have developed a novel method to incorporate reference genome alignment information by building loci *de novo* and subsequently integrating the alignments of the consensus sequences into their *de novo* dataset (available in STACKS 1.42).

To extract meaningful biological information from a RAD-seq dataset, it is crucial to explore the parameter space and assess how the analysis software interacts with the biological signal. Below, we use empirical datasets from three different

Table 2. Decision framework for each main STACKS parameter that control locus formation and polymorphism in STACKS, the values that users should test and considerations when exploring the parameter space

Parameter	Test	Decision framework			Other considerations
m	3–7	↑ if coverage <15×	↑ if contamination	↑ if conducting phylogenetics	if >m6 disable use of secondary reads
M	1–8	↑ if high polymorphism	↑ if high genomic divergence	↓ if repetitive or polyploid genome	if *M* is too high, paralogous loci can be filtered in populations. Rescale parameters with increased read length (250 bp+)
n	= *M* = *M* + 1 = *M* − 1	↑ if high polymorphism	↓ if sampled from same population	↑ if conducting phylogenetics	plot to observe changes in SNP heterozygosity and fixation

species to demonstrate (i) a method of *de novo* parameter optimisation, and (ii) how integrating alignment information from a reference genome can be used to supplement alignment information to loci built *de novo*.

Materials and methods

RAD-SEQ DATASETS

We used three empirical RAD-seq datasets for analysis, representing a phylogenetically diverse group of organisms (Table 3). The first dataset (*TRT*) consists of data from brown trout (*Salmo trutta* L.) samples occupying two different environmental niches (Paris, King & Stevens 2015): clean (8 individuals) and metal-impacted (8 individuals) sites. The second dataset (*PGN*) is from the king penguin (*Aptenodytes patagonicus*), from two different colonies on different archipelagos (Cristofari *et al*. 2016a): KER (8 individuals) and PCM (8 individuals). The final dataset (*ETW*) contains 16 red earthworm (*Lumbricus rubellus*) individuals from a single population (OL2; Giska, Sechi & Babik 2015).

Analyses of the datasets were performed using STACKS version 1.42 (Catchen *et al*. 2011, 2013a). Data quality was first checked using FastQC (Andrews 2010). If required, reads were cleaned and demultiplexed using the `process_radtags` program. Using each dataset's respective reference genome, the restriction enzyme used and type of RAD library preparation, we estimated the expected number of cut sites and RAD loci *in silico* (Table 3). For *ETW*, two estimates were made; the first by searching for the *Mse*I cut site 200–400 bp downstream/upstream of *Sph*I (liberal estimate) and a second estimation accounting for a second (*Mse*I) cut site occurring within 0–200 bp downstream/upstream of the first cut site (representing allele dropout; conservative estimate). For each dataset, we used `kmer_filter` in STACKS to visualise the error profiles of the cleaned RAD-seq reads; resulting K-mer frequency distributions were plotted in Gnuplot (version 5.0, http://gnuplot.info).

DE NOVO MAP AND PARAMETER OPTIMISATION

For each dataset, we ran *de novo map* several times, varying just one parameter with each parse of the program. For the primary analysis, we varied the `ustacks` *m* parameter from 1 to 6 (m1–m6), the `ustacks` *M* parameter from 0 to 8 (M0–M8) and the `cstacks` *n* parameter from 0 to 10 (n0–n10), while keeping all other parameters consistent (m3, M2 and n0). For further validation, we repeated these same runs with the defaults set to (m6, M4/M6, n0).

We extracted and collated data on *de novo map* assembly metrics for each parameter iteration including: (i) the number of assembled loci; (ii) the number of polymorphic loci and (iii) the number of SNPs for the parameters *m* and *M*. For the *m* parameter, we also collected data on coverage. This information is reported by STACKS' component programs and captured in the log files of both *de novo map* and *ref map*. We extracted these data using simple shell scripts, which we provide (distributed with STACKS).

Differences in both natural polymorphism and read depths can vary across individuals (Davey *et al*. 2013), and so exploring discrepancies between the individuals across a dataset are important. For each parameter run for *m*, *M* and *n* we visualised the data across the population of samples using STACKS' `populations` module, varying the value for the *r* parameter so that a locus had to be present in a minimum of 40%, 60% and 80% of individuals (for *m* and *M*) and 80% of individuals for *n*. To further interrogate how *M* controls polymorphism, we assessed how many new polymorphic loci were identified across 80% of the population (*r80 loci*) for each increment of *M*.

To observe variation in changing the number of fixed differences allowed between loci across individuals when forming the *catalog* (the *n* parameter), we compared the SNPs from each individual with those in the *catalog* using a custom Python script (*count_fixed_catalog_snps.py*; distributed with STACKS). The script tabulates: (i) the number of heterozygous SNPs found in each individual; (ii) the variable sites identified across the whole population; (iii) the number of SNPs found in the *catalog*; and (iv) the number of SNPs found in the *catalog* but not found in any individual sample – that is, the fixed SNPs captured during the construction of the *catalog*. For increments of *n*, we also calculated the distribution of SNPs per *catalog* locus, as well as SNP distributions across polymorphic loci identified in 80% of the population (*r80 loci*).

Metrics for *m*, *M*, and *n* were plotted in GNUPLOT (version 5.0), visually assessed and optimal sets of parameters were chosen, specific to each dataset. The results generated by the optimal parameters were plotted against those obtained from an analysis using the STACKS default parameters (m3, M2, n1), and common population genetic statistics were calculated to assess how the optimal parameter runs compared to using the default values.

COMPARING LOCI BUILT *DE NOVO* TO REFERENCE-ALIGNED RAW READS

Our next objective was to compare loci derived from reference-aligned raw reads (*ref map*) with loci first assembled *de novo* and then aligned back to a genome – integrating the alignment position for each locus back into the STACKS output files (*integrated*).

Table 3. Details of the three Restriction site-Associated DNA sequencing datasets used for analysis

Dataset	Species	Estimated genome size	RAD flavour	Restriction enzyme	Reference genome	Reference estimation of cut sites	Reference estimation of RAD loci	*De novo* estimation of RAD loci
TRT	Brown trout (*Salmo trutta*)	3 Gb	Single digest (1 × 100 bp)	*Sbf*I	Atlantic salmon (*Salmo salar*)	58 197	116 394	257 322 (62 767-82 406)
PGN	King penguin (*Aptenodytes patagonicus*)	1·2 Gb	Single digest (2 × 100 bp)	*Sbf*I	Emperor penguin (*Aptenodytes forsteri*)	43 435	86 870	156 725 (20 664-71 122)
ETW	Red earthworm (*Lumbricus rubellus*)	420 Mb	Double digest (1 × 100 bp)	*Sph*I and *Mse*I	Red earthworm (*L. rubellus*)	127 402/278 544	143 380/399 731	832 898 (5884-56 666)

For each of the three species: the dataset abbreviation used; estimated genome size of the species; the type of RAD sequencing and restriction enzyme(s) used; the reference genome used for alignment analysis; *in silico* estimation of the number of RAD cut sites and resultant RAD loci using the reference genome and the number of RAD loci assembled *de novo* using optimal parameters. For *ETW*, two estimates of reference-related RAD cut sites and loci are presented to show both conservative and liberal estimates based on ddRAD digest sites and allele dropout. For *de novo* estimation, the total number assembled across the dataset is presented, and numbers in brackets represent those assembled across 40% and 80% of individuals within the dataset.

For the *ref map* analyses, clean raw reads were aligned to either a complementary reference genome of the same species or to that of a closely related sister species (Table 3). Raw reads were aligned using GSNAP (version 2015-12-31; Wu & Nacu 2010) specifying a maximum of five mismatches (-m 5), an indel penalty (-i 2) and turning off terminal alignments (–min-coverage = 0·95). Only reads that aligned uniquely (`unpaired_uniq`) were used. *TRT* was aligned to the closely related Atlantic salmon genome (*Salmo salar*, NCBI accession GCA_000233375.4), *PGN* was aligned against the emperor penguin (*Aptenodytes forsteri*, NCBI accession GCF_000699145.1), a close relative of the king penguin, and *ETW* was aligned against a draft genome of the same species (P. Kille and L. Cunha, Cardiff University). Resulting alignment files were run through *ref map*.

For the *integrated* analyses, we used GSNAP (using the same parameters as outlined above) to align the consensus sequence of the catalog loci from the optimal *de novo map* runs for each species against their respective genomes and used a Python script (*integrate_alignments.py*; distributed with STACKS) to integrate alignment information back into the original *de novo map* output files. We compared the number of uniquely mapping loci that were assembled using an alignment from *ref map,* to those aligned using *de novo map* consensus sequences – the *integrated* method.

Results

RAW DATA STATISTICS

Reads from the *TRT* samples were 95 bp in length with an average of 2 880 327 (± 290 590 SE) reads per sample (Table S1, Supporting Information); the *PGN* dataset was 96 bp with 3 312 159 (± 337 146 SE) reads per sample; the *ETW* dataset was 95 bp with 3 629 617 (± 409 701 SE) reads per sample. To get a sense of error and repeats, we generated the distribution of K-mer counts from our set of raw reads. Each true RAD locus can be described by a set of K-mers (subsequences of length *K*) and those K-mers will appear in the distribution in proportion to the number of times a RAD locus was sequenced. The mean of the K-mer distribution reflects the sequencing depth, rare K-mers represent sequencing or PCR error and frequent K-mers describe repetitive loci. Figure S1 demonstrates that *ETW* contains significantly more error than *TRT* and *PGN*.

CHOOSING A VALUE FOR THE MINIMUM NUMBER OF RAW READS (*M*) REQUIRES MEDIATING THE PROMOTION OF ERROR READS TO PUTATIVE ALLELES VS. EXCLUDING TRUE ALLELES

The first stage in the assembly of loci *de novo* (Catchen *et al.* 2011) is to collapse identical raw reads into STACKS and consider them putative alleles. The number of raw reads required to form an allele is governed by *m*. The general pattern for *m* is that using higher values increased average sample coverage (Fig. 1), but decreased the number of assembled loci (Fig. 2a) and the amount of polymorphism (Fig. 2b,c).

Figure 1 (green) shows that stack coverage improved with increasing values of *m*, and that after merging putative alleles into loci, coverage further increased across all datasets (Fig. 1, purple). Coverage between the datasets varied with *PGN*

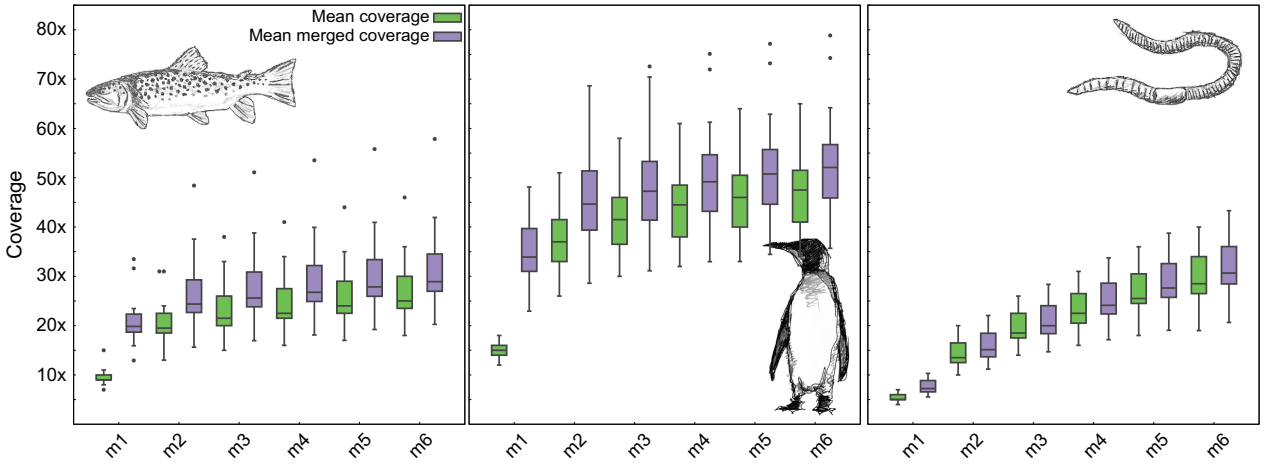


Fig. 1. Plots of mean coverage for *TRT, PGN* and *ETW*. Coverage for each of the 16 individuals is represented as a box plot, where mean coverage (in green) is the average mean coverage for primary reads and mean merged coverage (in purple) is the coverage after merging alleles.

showing the highest coverage, followed by *TRT* and finally *ETW* (Table S2). With the STACKS default value of m3; the average coverage for *TRT* was 23× (17–38×). Merging putative alleles into loci increased coverage further to 28× (16–51×). *PGN* had the highest coverage: 42× (30–58×); and a merged coverage of 49× (31–72×). *ETW* had the lowest coverage at 19× (14–23×), and only reached 21× (15–28×) after merging alleles.

At a value of m1, every raw read is treated as a putative allele and thus all datasets showed the most loci assembled for this value (Fig. 2a; Table S3a). Proportionately, we saw the fewest number of polymorphic loci and the fewest number of SNPs (Fig. 2b,c; Table S3b,c) for m1. Moreover, almost none of these SNPs was shared across the population.

When we increased m1 to m2, the average number of loci formed dropped dramatically. For *ETW*, ~50% more loci were assembled at m1 compared to m2, suggesting that *ETW* contains many unique reads, potentially indicating a large amount of PCR or sequencing error. This was confirmed in the K-mer distribution plots (Fig. S1). In all datasets, we observed a large drop in the number of loci assembled between m2 and m3, but after m3 the number of loci that were built stabilised.

The average number of assembled loci was significantly higher in *ETW* (138 905 at default m3) despite having the lowest coverage, moderate in *TRT* (92 626, default m3) and the fewest loci were assembled in *PGN* (58 876, default m3), despite its very high coverage. These results are consistent with our *in silico* estimates of the number of RAD loci (Table 3). The *TRT* dataset contained the lowest variance and also showed high consistency in the homologous loci being repeatedly assembled in 60% and 80% of the population. *PGN* showed a high, but consistent variance in the number of loci assembled, with many fewer loci shared across 60% and 80% of individuals in the population; however, the high coverage and consistent results with m3–m6 imply a true biological signal, and the distribution of assembled loci was likely bimodal. Indeed, considerably more loci were assembled in the KER population (m3: 74 646 ± 6934), compared to the PCM population (m3: 49 280 ± 5363), also representing higher levels of polymorphism (m3: KER: 7717 ± 1146; PCM: 2475 ± 427). Such patterns could be the result of a batch bias generated by sequencing the two populations on different lanes. Alternatively, this could be due to biological processes such as population bottlenecks, founder effects or stochastic demographic histories. Conversely, *ETW* showed very few loci assembled consistently across the population; combined with the low coverage, this again suggests a signal of error.

The average number of polymorphic loci in *TRT* and *PGN* (Fig. 2b) decreased by ~100 loci with each higher value of *m*, suggesting that increasing read depth excluded only a small amount of biological polymorphism. Given that *M* was fixed for these tests, we determined that *m* does not have a significant effect on polymorphism or SNP detection (Fig. 2c) and this was further confirmed in additional parameter tests (Fig. 2c; Tables S4 and S5). In *ETW,* several hundred loci were lost with each increasing value of *m*, corroborating the hypothesis for a high amount of error.

In a biologically unambiguous dataset with reasonable sequencing coverage (e.g. *TRT*), the *m* parameter converges after a value of 3 and on its own does not have a large impact on the detection of polymorphism. However, by varying *m*, we observed differences in a high coverage dataset that appeared to have different subsets of loci in different individuals (*PGN*) vs. a dataset with low coverage and apparently high error that does not show a strong biological signal (*ETW*).

SETTING VALUES FOR THE MAXIMUM NUMBER OF MISMATCHES (*M*) IS SPECIES SPECIFIC AND REQUIRES A BALANCE BETWEEN UNDERMERGING AND OVERMERGING LOCI

After putative alleles are formed, STACKS performs a search to match alleles together into putative loci. This search is governed by the *M* parameter, which controls for the maximum number of mismatches allowed between putative alleles. The general pattern for *M* was that higher values decreased the

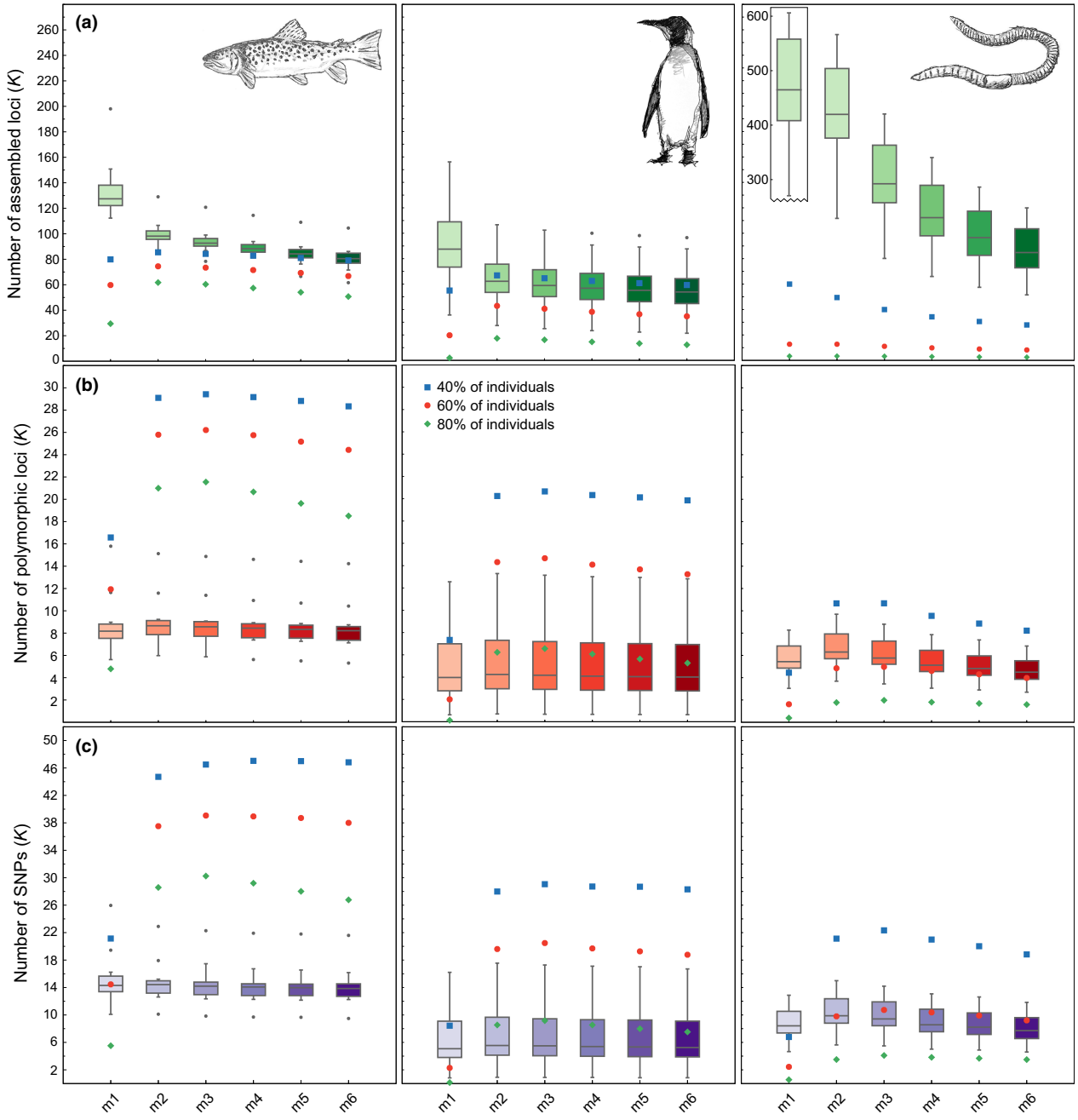


Fig. 2. Plots of iterating values the minimum number of raw reads required to form a stack (*m*) for the metrics: (a) the number of assembled loci; (b) the number of polymorphic loci and (c) the number of SNPs in *TRT, PGN* and *ETW*. Blue squares represent data found in at least 40% of the samples, red circles 60% and green diamonds 80% (*r80*).

number of assembled loci (Fig. 3a), but increased the number of polymorphic loci (Fig. 3b) and the number of SNPs (Fig. 3c).

Correctly setting *M* requires a balance – set it too low and alleles from the same locus will not collapse, set it too high and paralogous or repetitive loci will incorrectly merge together. When alleles from the same locus are undermerged, the software will incorrectly consider them as independent loci. When loci are overmerged because they happen to be close in sequence space, an errant locus with false polymorphism will result. Therefore, *M* is particularly dataset-specific because it depends on the natural levels of polymorphism in the species, as well as the amount of error generated during the preparation and sequencing of the RAD-seq libraries.

The number of assembled loci decreased with increasing values for *M* (Fig. 3a). For *TRT*, the proportion of putative alleles that collapsed levelled out between M3 and M4 (Fig. 3a; Table S6a). In *PGN* these patterns started at lower mismatch values, between M1 and M2. In *ETW*, the numbers that collapsed began to plateau between M2 and M3.

For the number of polymorphic loci (Fig. 3b; Table S6b) and the number of SNPs (Fig. 3c; Table S6c), obviously M0 is

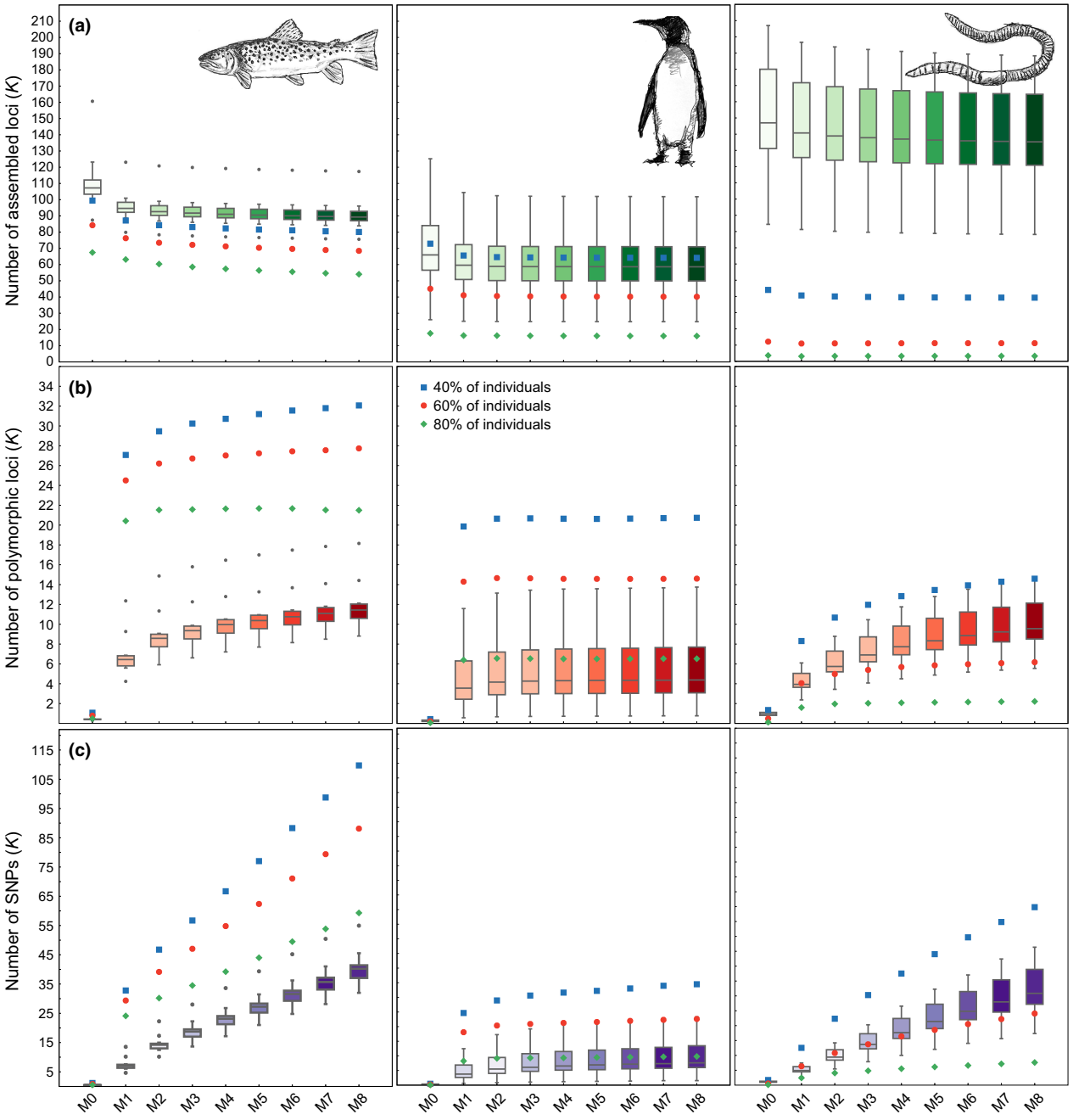


Fig. 3. Plots of iterating values for the distance allowed between two stacks (*M*), for the metrics: (a) the number of assembled loci; (b) the number of polymorphic loci and (c) the number of SNPs in *TRT, PGN* and *ETW*. Blue squares represent data found in at least 40% of the population samples, red circles 60% and green diamonds 80% (*r80*).

incorrect. A high number of loci was assembled (Fig. 3a) simply because putative alleles were not correctly merged into heterogeneous loci. Increasing *M* from 0 to 1 permitted alleles with a single polymorphism to merge and with increasing values of *M* no clear limit was seen in the average amount of polymorphism detected.

At M1, the majority of polymorphism and SNPs were already captured in *PGN*, and the amount of polymorphic loci was relatively uniform; the highest polymorphism detected across 80% of the population was at M2. In *TRT,* the average number of polymorphic loci identified within individuals began to flatten out at M5, and M5 also provided the highest amount of polymorphism across 80% of the population. Alternatively, in *ETW*, although polymorphic loci within individuals began to plateau at approximately M3, polymorphism at 80% continued to increase with incremental values of *M* (even up to unrealistically high levels of M11; Table S7), suggesting that the dataset contained a small number of loci with a high-density of SNPs.

In both *TRT* and *ETW* (Fig. 3c) we noticed a steep increase in the number of SNPs obtained with increasing *M*; however, a much smaller increase was detected in *PGN*. The steeper increase in SNPs in *TRT* reflects the higher value of *M* required

to detect the maximum amount of polymorphism across 80% of the population. Although *TRT* was quite polymorphic, a large proportion of these polymorphisms was not shared across the population, and many of the alleles were at low frequency.

When we observed the number of assembled loci (Fig. 3a) together with the number of polymorphic loci (Fig. 3b) and the number of SNPs called (Fig. 3c), we saw a small number of loci that contained many SNPs added with each increment of *M*. This is also reflected in the haplotype diversity in Fig. S2. With high coverage, these are likely to be true loci as the SNP calling algorithm can assess SNP calls with a higher likelihood. However, with low coverage these are more likely to be erroneously assembled loci.

Figure 4 describes how increasing values for *M* contributed new broadly shared, and therefore likely real, polymorphic loci (i.e. loci found in 80% of the population or, *r80 loci*). The number of novel polymorphic loci peaked at M3/M4 for *TRT*, M1/M2 for *PGN* and M2/M3 for *ETW*, and importantly, after a particular value for *M*, identification of new polymorphic loci appears to be robust to any further increases in *M* (Table S8). The data here corroborate the levels of polymorphism displayed in Fig. 3b,c in illustrating the relative homogeneity of *PGN*, compared to *TRT* and *ETW* which were more polymorphic and required higher *M* values.

SELECTING A VALUE FOR THE NUMBER OF MISMATCHES IN THE *CATALOG* (*N*) REQUIRES ASSESSING THE MAXIMUM AMOUNT OF 80% POLYMORPHIC LOCI (*R80*) AROUND *N* = *M*

After loci are assembled in each individual sample, they are compared across samples to match homologous loci into a single *catalog* locus for the population. The *n* parameter controls for the number of mismatches, or fixed differences, allowed during this process. The general pattern for *n* is that higher values increased the number of fixed differences found between samples (Fig. S3; Table S9). Choosing the value for *n* involves a trade-off between setting it too low and failing to find homologous loci in different samples that contain fixed differences, and setting *n* too high and thereby further collapsing loci close together in sequence space within and across samples.

At n0, some loci (and the variants they contained) had not been integrated into the *catalog*. However, once *n* was set higher than 0, the number of variant sites in the population stabilised. In all three datasets, increasing values of *n* provided a linear increase in the identification of more fixed differences between individuals, and increasing values of *n* did not cause the number of fixed differences to plateau (Fig. S3). There appears to be an unlimited number of loci that can be connected together in the *catalog* with increasing values of *n*, so we chose to focus again on optimising for broadly available loci, *r80 loci*.

By iterating over values for n0-n10 for M2, M4 and M6, we saw that at M2 all datasets contained the highest amount of *r80 loci* at $n = M$. At M4 and M6, the highest *r80* values were obtained for values of *n* one iteration either side of $n = M$, so that $n = M-1$ or $n = M + 1$ (Table S9). The optimal value of *n* for *TRT* (n4) provided a total of ~153K SNPs, of which ~24K (15%) were fixed (Table S10). In *PGN*, n2 provided ~62K SNPs, with ~8K being fixed (13%) and for *ETW* (n3) contributed ~156K SNPs and ~33K (21%) of these SNPs were fixed.

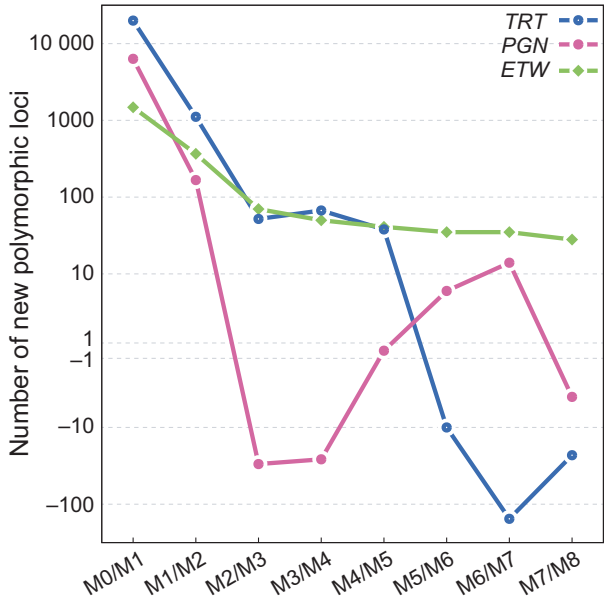


Fig. 4. Plots of the number of new polymorphic loci (*r80 loci*) added for each iteration of *M* (the distance between stacks) for the three datasets: *TRT*, *PGN* and *ETW*.

THE MAXIMS FOR SELECTING OPTIMAL PARAMETERS IN STACKS

Through collecting and plotting metrics of the iterations of *m*, *M* and *n* we have identified two general rules that allow for the identification of optimal parameters: (i) the 80% polymorphic (*r80*) loci rule, and (ii) $n = M$ plus or minus one iteration for linking loci together across samples. There are many possible metrics related to the *de novo* construction of loci we could focus on to optimise parameters. We chose to select a stable set of loci that are highly replicated across the population; these *r80* loci are unlikely to be derived from paralogous or repetitive sequence, or have a lot of sequencing error, and serve as a proxy for the true genome. Importantly, although we are using the *r80* loci as an optimisation target, we are not required to use only the *r80* loci for downstream analyses – we still have all subsets of the loci available to us, we have simply used the *r80* loci to optimise the assembly of all loci.

A value of m3 was optimal in providing the highest amount of polymorphism across all three datasets at 40%, 60% and most importantly 80% of the population. We observed that the *PGN* dataset was relatively monomorphic, and the highest amount of polymorphism identified at *r80* was detected at a value of M2. The *TRT* and *ETW* datasets showed higher polymorphism; indeed, the largest amount identified at *r80* was

found at M5 for *TRT*, making this the most polymorphic dataset. In *ETW*, the polymorphism at *r80* continued to increase with higher *M* values, but Fig. 4 showed that very few new polymorphic loci were added at *M* values greater than 3. Given the potential error in *ETW*, for this dataset, M3 also provides a balance between obtaining true polymorphism and introducing sequencing error.

All datasets continued to show more fixed SNPs with increasing values of *n* (Fig. S3). *PGN* showed a very gentle increase, *TRT* a moderate increase and *ETW* the steepest increase in fixed SNPs. This alone, however, does not provide enough information to choose a value for *n*. We found that the highest polymorphism of *r80 loci* resulted from $n = M$, $n = M-1$ or $n = M + 1$ (Table S9) and suggest this as the best method for obtaining the optimal value for *n*. The following optimal parameters were therefore chosen so that: *TRT* = m3, M5, n4; *PGN* = m3, M2, n2; *ETW*: m3, M3, n3.

Using these optimal parameter sets, in *TRT*, ~244K loci were identified across the population, 22% were polymorphic containing an average of 3·18 SNPs per locus; for *PGN*, ~155K loci were assembled, 32% were polymorphic, with 1·74 SNPs per locus; in *ETW*, ~865K loci were assembled, 13% were polymorphic, with 2·83 SNPs per locus. We suggest assessing and reporting these metrics in any RAD-seq analysis.

Figure 5 shows the main metrics plotted for the STACKS default parameters (m3, M2 and n1) and the optimal parameters for each dataset. Significantly, when compared to the defaults, the optimal parameters modified the number of constructed loci, the polymorphism levels and the numbers of SNPs (Fig. 5, Table S11). Moreover, the number of polymorphic loci discovered across 80% of the population (*r80*) increased with optimal parameters (Fig. 5, green dots). Although overall measures of population genetic statistics were not considerably altered (Table S12), changing the values for *M* considerably altered global SNP and haplotype diversity statistics across all three datasets (Fig. S2, Table S13).

DE NOVO ANALYSIS RECOVERS MORE LOCI COMPARED TO REFERENCE-ALIGNED RAW READS

It was apparent that in all three datasets the *de novo-integrated* approach recovered thousands more loci than *ref map* (Fig. S4; Table S14). In *TRT*, a total of ~116K *de novo* consensus loci aligned to the reference genome. Of this total, ~84K represented consensus loci that aligned to unique positions in the reference. This compares to just ~66K alignments using *ref map*. Similarly, in *PGN* over 117K consensus loci aligned using the *integrated* method, ~104K of which were aligned uniquely, in comparison to under 92K using *ref map*. The most obvious discrepancy was seen with the *ETW* dataset where ~253K loci aligned with the *integrated* method (~172K unique alignments) in comparison to just ~100K loci using *ref map*. In the large majority of cases, raw reads that were poorly aligned compared to well-integrated consensus *de novo* sequences were due to either insertions into the reference, or deletions in the sequence relative to the reference. Our comparison is of course dependent on the corresponding reference alignment parameters used; more promiscuous values may have closed the gap between *de novo map* and *ref map* a small amount.

Discussion

Restriction site-Associated DNA sequencing data interrogation is a non-trivial, yet necessary process for avoiding inaccurate building of loci, erroneous SNP calls and a loss of important biological information. Our aim was to develop a straightforward method for surveying and visualising the trends of assembly metrics from the output of STACKS, and we suggest that by doing so, researchers can accurately identify optimal parameter sets. We present two general maxims that can be used to optimise parameter space, but highlight that the *r80* rule, which uses a set of polymorphic loci repeatedly assembled across a set of samples, will provide an effective

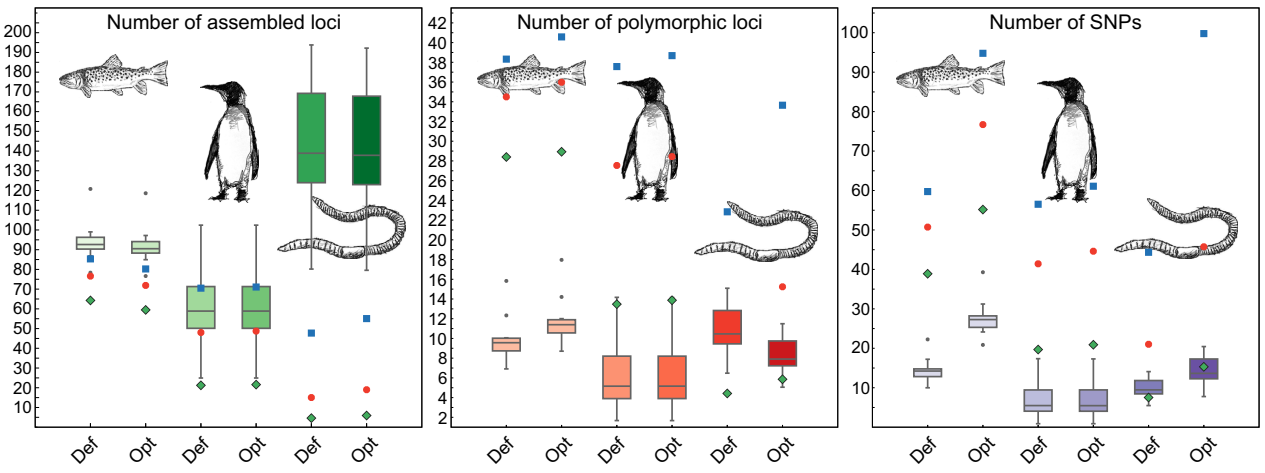


Fig. 5. Plots showing the differences between the default STACKS parameters (m3, M2, n1) and the optimal parameters selected for each dataset: *TRT:* m3, M5, n4; *PGN:* m3, M2, n2 and *ETW:* m3, M3, m3 for the number of assembled loci, the number of polymorphic loci and the number of SNPs. For every dataset, the optimal parameter sets gave the highest amount of polymorphic loci in 80% of the population (*r80 loci*, green diamonds).

optimisation target. Furthermore, we have shown that incorporating alignment information from a reference genome is an advantageous approach compared to using a standard method of alignment, particularly if using a draft reference genome or if the genome is phylogenetically distant from the focal species.

The STACKS *de novo* assembly algorithm proceeds in each individual sample in several stages (Catchen *et al*. 2011). First, exactly matching reads are collapsed into putative alleles, controlled by *m*. Sets of exactly matching reads with a size smaller than *m* are set aside and are referred to as secondary reads. Putative alleles are then compared against one another and collapsed into putative loci, controlled by *M*. Once the loci are formed, the secondary reads are aligned back against assembled loci with a relaxed mismatch limit (the `-N` parameter to `ustacks`) to increase locus depth. Next, the SNP model is executed and it evaluates the evidence at each nucleotide position for a homozygote or heterozygote. The model takes into account depth of coverage and the amount of error present (i.e. non-matching nucleotides at the current position) in a maximum likelihood statistical framework. Finally, after loci are assembled in each sample, homologous loci must be matched across samples to make a population-wide *catalog* of loci. Loci are compared across samples and sequences within a certain distance are collapsed into a single *catalog* (population-wide) locus, governed by *n* (`cstacks`).

We recommend scrutinising sample coverage as the first assembly metric when assessing the suitability of any given RAD-seq dataset and we suggest aiming for coverage thresholds greater than 25×. Fountain *et al*. (2016) showed that genotyping error rates were considerably higher when coverage was between 5 and 10×, and error rates were mostly robust to variation in sequence quality when coverage was ≥10. However, published datasets with coverage below 10× are not uncommon (e.g. Xu *et al*. 2014; Boehm *et al*. 2015; Ivy *et al*. 2016; Razkin *et al*. 2016).

As seen here, the main way the *de novo* assembly algorithm interacts with coverage is through the *m* parameter. If *m* is set to 1, each raw read becomes a putative allele and there are no secondary reads. However, as *m* is increased the status of a putative allele requires more evidence and the number of secondary reads grows. If *m* gets too large, alleles lacking coverage will not be recognised and will become secondary reads. These secondary reads will still be made available to the SNP calling model, as long as one of the two alleles was assembled, but if both alleles fall below a high *m* threshold the locus will be lost.

In the literature, wide ranges of values for the *m* parameter have been used. In some cases, the investigation was specifically addressing phylogenetic differences between divergent species, and hence underdetection of species-specific polymorphism when building loci was inconsequential (e.g. m50 – Keller *et al*. 2012; m125 – Wagner *et al*. 2013). Furthermore, a higher value of *m* may be required to exclude exogenous contamination (Trucchi *et al*. 2016). However, other studies have used inappropriately high values of *m* (Jezkova *et al*. 2015; Palaiokostas *et al*. 2015; Suyama & Matsuki 2015), presumably due to the assumption that higher values will reduce genotyping error. If *m* is set unrealistically high, secondary read incorporation must be disabled (`-N` option). This was demonstrated when comparing m3 and m6 with iterative values for *M*, where at M0 more SNPs were supposedly called at m6, compared to m3 (Table S5). However, these are not true SNPs, but simply represent sequencing error incorporated by secondary reads at higher read depths.

We recommend setting *m* high enough only to deny errant reads the status of a putative allele. We therefore do not suggest using a value for $m > 5$ and we have demonstrated here that the default value of m3 was favourable for all test datasets. Testing a range from 3 to 7 should allow the correct exploration of this parameter under most biological scenarios. If coverage is exceptionally high (i.e. >40×), or levels of polymorphism are exceptionally low (as in the *PGN* dataset), *m* can still be left at a moderate value, but secondary reads can be discarded entirely (setting `N` to 0), making for a clearer signal (e.g. Longo & Bernardi 2015). Despite controlling for the minimum read depth of alleles at the `ustacks` phase of the pipeline, many studies also incorporate a minimum stack depth required for individuals at a locus in the `populations` module of STACKS (e.g. Gaither *et al*. 2015; Ivy *et al*. 2016; Kjeldsen *et al*. 2016). Such a method is undesirable, as read depth has already been accounted for by the SNP model. Once the SNP model has made a determination, its evaluation should be trusted and using further non-statistically based limits on depth of coverage is ill advised and will result in the arbitrary dropping of loci.

Species with higher levels of polymorphism will require higher values of *M* (Campagna *et al*. 2015; Ravinet *et al*. 2016), as do studies assessing levels of genomic divergence (Jones *et al*. 2013; Lozier *et al*. 2016). If *M* is too small, alleles will not collapse into loci within individuals. When the *catalog* is constructed, different individuals will map alternative alleles together creating loci that appear homozygous, while the other alleles are excluded and may themselves form distinct loci. On the other hand, if *M* is too high, repetitive or paralogous loci will be erroneously merged together. These loci are noticeable, as they will appear as heterozygous in a large majority of individuals; they can be filtered out by the `populations` module after the main pipeline runs, based on their high heterozygosity (`--max_obs_het` option).

The three datasets examined here required different values for *M*. The *ETW* dataset is likely quite polymorphic; high levels of polymorphism have been shown in the *Lumbricus* genus (Kautenburger 2006; Shepeleva *et al*. 2008; Donnelly *et al*. 2014). However, the low coverage and potential error in the dataset means a definitive determination of true polymorphism could not be reliably assessed, but M3 provided the highest amount of *r80 loci*.

For the *TRT* dataset, M5 was the most suitable value for identifying polymorphism, and this is considerably higher than M2 as most commonly used in salmonid RAD-seq studies (Lemay & Russello 2015; Bernatchez *et al*. 2016) and in other teleost species (Catchen *et al*. 2013b; Martin & Feinstein 2014; Fowler & Buonaccorsi 2016). Observations of the *TRT* dataset suggested that a large proportion of the alleles were at a low

frequency; these could be filtered out after the main pipeline executes by implementing a minor allele frequency threshold (`--min_maf` option to `populations`).

On the other hand, the *PGN* dataset was relatively uniform and exhibited low levels of polymorphism. This corroborates with limited genetic variation and high gene flow that has been shown to exist in penguin colonies (Roeder *et al*. 2001; Nims *et al*. 2008; Freer *et al*. 2015). The bimodal distribution for the number of loci observed in *PGN* corroborates well with the known stochastic demographic history of the species (Cristofari *et al*. 2016a). Considerably fewer loci were observed in the PCM population, a colony sampled from an archipelago that was recolonised later after glaciation, that is, later than the archipelago home to the KER colony. While the length of the loci that were generated should be considered, with most short-read sequencing (~100 bp) surveying M1 to M8 should allow sufficient mismatch parameter space exploration for the large majority of RAD-seq datasets. Of course, if read lengths are longer (250 bp+), users should rescale the parameter, testing higher values of M as appropriate for the longer read length.

A main consideration when deciding on the number of mismatches allowed in the *catalog* (n) is how many fixed differences might be expected between individuals. In a closely related set of populations, it makes intuitive sense to set $n = M$, as M controls matching alleles within an individual, and n controls matching of the same set of alleles across individuals. If, by chance for a particular locus, only homozygous individuals were sampled, setting $n = M$ will allow the correct relationship for that locus to be recovered across individuals. If population sample sizes are large (so heterozygotes should have been found), and fixed differences are rare, it may make sense to set n to a value less than M (Barnard-Kubow, Debban & Galloway 2015; Saenz-Agudelo *et al*. 2015). Alternatively, if the samples originate from highly divergent individuals (Ravinet *et al*. 2016; Rougemont *et al*. 2016), or phylogenetic relationships are being explored between species (Combosch & Vollmer 2015; Tariel, Longo & Bernardi 2016), then a higher value of n may be required to detect these fixed polymorphisms. This being said, it may be difficult to derive a biological judgement of the known amount of differentiation between individuals – for example, if cryptic population structure exists. Therefore, it is essential to plot and explore how iterative values for this parameter affect how many fixed differences are detected.

Although the number of fixed SNPs continued to grow with incremental values for n for both *TRT* and *ETW*, the highest amount of *r80 loci* was equal to $n = M$ for *ETW*. High levels of intraspecific divergence are common in earthworm species (Klarica *et al*. 2012) and, in particular, natural populations of *L. rubellus* have been shown to consist of highly divergent mtDNA lineages representing a complex of cryptic species (King, Tibble & Symondson 2008). Thus, high divergence between individuals, even within the same population is very likely and validates the patterns we observed. The *TRT* dataset showed a moderate number of fixed SNPs, and the highest number of polymorphic *r80 loci* was recovered for $n = M-1$. This is in accordance with high levels of differentiation between geographically proximate populations of the species, occupying metal-contaminated and clean rivers (Paris, King & Stevens 2015). In king penguins, studies have shown low levels of genetic divergence both within (Cristofari *et al*. 2015) and between (Freer *et al*. 2015) colonies, and it has recently been suggested that Antarctic penguins should be considered single panmictic populations due to extreme levels of genetic homogeneity and low F_{ST} values between populations (Cristofari *et al*. 2016b). Coupled with the low levels of within-individual polymorphism, $n = M$ resolved the most polymorphic *r80 loci* across the populations for this dataset. Thus, the general maxims and visualisation of the RAD-seq plots created here and the subsequent parameters chosen corroborate well with the known biology of these species.

Our data clearly show (Fig. S5; Table S15) that in all three datasets there are a small number of loci containing high densities of SNPs. These loci appear in a *de novo* dataset with high values of M, and also appear in a reference-aligned dataset with promiscuous alignment parameters. Including these loci will affect standard population genetic statistics (Fig. S2; Table S13), naturally inflating them. Experiments that rely only on downstream population genetic measures to assess accuracy need to account for this (Rodríguez-Ezpeleta *et al*. 2016; Shafer *et al*. 2016) or they may produce misleading results.

A reference genome is a great asset in a RAD-seq analysis, enabling loci to be considered positionally and statistics to be computed across chromosomes. However, one aspect that may be both an asset and a liability is that the reference genome can act like a strict filter, and so what a researcher cannot observe as aligned on a chromosome is not included. In addition, alignment programs align each read independently and alignment algorithms can score different combinations of gaps and mismatches the same way, which can lead to alignment variation. Furthermore, RAD-seq is a key technology for non-model organisms, so, often a reference genome, if available, is a draft, or is a closely (or not so closely) related species, exacerbating the above problems. By building loci *de novo*, individual reads have already been merged into biologically plausible loci, thereby leveraging the maximum amount of information obtained from each individual read. In a straight alignment approach, each read is treated independently of every other read. Our method of building loci *de novo*, aligning the consensus sequences to a reference and then reintegrating alignment positions into the *de novo* data to compute chromosome-level statistics combines positive aspects of both *de novo* and reference analyses.

Other studies have used a version of the *de novo-integrated* approach (Jones *et al*. 2013; Wagner *et al*. 2013). In all datasets analysed here, the *de novo* and *integrated* method performed better than alignment and *ref map*. In both the brown trout (*TRT*) and king penguin (*PGN*) cases, the reference genome was not from the same species, but belonged to the same genus, suggesting that the genome of a closely related species can be utilised. Gene synteny is conserved between brown trout (*S. trutta*) and Atlantic salmon (*S.* salar) (Gharbi *et al*. 2006)

and the species are known to hybridise (Leaniz & Verspoor 1989; Elo *et al.* 1997). Similarly, phylogenetic relationships within the genus *Aptenodytes*, i.e. the king penguin (*A. patagonicus*) and emperor penguin (*A. forsteri*), are closely related (Baker *et al.* 2006; Ksepka, Bertelli & Giannini 2006). In both cases, the RAD-seq reads for the *TRT* and the *PGN* datasets aligned successfully to the reference genome, but this was improved by building loci *de novo*, most likely the result of constructing species-specific loci first, which accounts for the variance between the *reference* genome and the samples used.

In contrast, the genome used for the red earthworm *ETW*, *L. rubellus*, was from the same species. The highly divergent *L. rubellus* species consists of 11 lineages (Sechi 2013); the samples used for the RAD-seq analysis were identified as lineages A, C and E (Giska, Sechi & Babik 2015) and the reference genome was sequenced from lineage B. COI analysis has shown that other lineages are >12% different to lineage B and only 20% of RNA-seq reads map from lineage A to lineage B (Kille, pers. comm., unpubl. data). The straight alignment method using *ref map* was relatively unsuccessful. However, here the *integrated* method successfully aligned all of the *de novo* reads to the genome.

We have demonstrated how the STACKS software can be tailored to a researchers' RAD-seq dataset; we have provided a method of parameter optimisation and exhibited how to effectively implement and test our optimisation strategy using STACKS. These parameters should be adjusted to complement the biology of the species being studied, the biological hypotheses, the library construction and error inherent in the dataset. While we focused on STACKS in this work, other RAD-seq analysis software uses the same algorithmic strategy, governed by analogous parameters. Analyses in other domains also use similar strategies, such as assembling metagenomic loci, or during alignment for shotgun resequencing analyses; thus, this basic approach has broad relevance and can be widely applied.

Authors' contributions

J.P. and J.S. conceived of the brown trout experiments, J.P. implemented the brown trout molecular work. J.P. and J.C. conceived and implemented the analytical methods. J.P., J.C., and J.S. wrote the manuscript.

Acknowledgements

This work was co-funded by the Environment Agency, Westcountry Rivers Trust and the University of Exeter. Overseas collaboration for the project was made possible by funding from The Genetics Society, Santander and the University of Exeter. Thank you to many RAD-seq workshop participants for invaluable insight and new ideas. We thank Dr Nicolas Rochette for his insights into parameter analysis. Thanks also to Dr Andy King for assistance with the brown trout data molecular work and analysis, and Guy Freeman and Martin Young for the species illustrations. Prof Peter Kille and Dr Luis Cunha, Cardiff School of Biosciences, Cardiff University, kindly provided the reference genome of *L. rubellus*.

Data accessibility

The brown trout samples (TRT) are archived at the NCBI Sequence Read Archive, BioProject PRJNA379215, accession numbers SRR5344602–SRR5344617. The king penguin samples (PGN) are archived at EBI-EMBL, Project ID 308448; run accession numbers SRR3177636–SRR317765, and the red earthworm samples (ETW) are archived at EBI-EMBL Project ID 296755; run accession numbers SRR2962225–SRR2962229 and SRR2962234–SRR2962244.

References

Amores, A., Catchen, J., Ferrara, A., Fontenot, Q. & Postlethwait, J.H. (2011) Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication. *Genetics*, **188**, 799–808.

Andrews, S. (2010) FastQC: a quality control tool for high throughput sequence data. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc (accessed 4 April 2017).

Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G. & Hohenlohe, P.A. (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, **17**, 81–92.

Arnold, B., Corbett-Detig, R.B., Hartl, D. & Bomblies, K. (2013) RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology*, **22**, 3179–3190.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A. & Johnson, E.A. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, **3**, e3376.

Baker, A.J., Pereira, S.L., Haddrath, O.P. & Edge, K.-A. (2006) Multiple gene evidence for expansion of extant penguins out of Antarctica due to global cooling. *Proceedings Biological sciences/The Royal Society*, **273**, 11–17.

Barnard-Kubow, K.B., Debban, C.L. & Galloway, L.F. (2015) Multiple glacial refugia lead to genetic structuring and the potential for reproductive isolation in a herbaceous plant. *American Journal of Botany*, **102**, 1842–1853.

Bernatchez, S., Laporte, M., Perrier, C., Sirois, P. & Bernatchez, L. (2016) Investigating genomic and phenotypic parallelism between piscivorous and planktivorous Lake Trout (*Salvelinus namaycush*) ecotypes by means of RADseq and morphometrics analyses. *Molecular Ecology*, **25**, 4773–4792.

Blanco-Bercial, L. & Bucklin, A. (2016) New view of population genetics of zooplankton: RAD-seq analysis reveals population structure of the North Atlantic planktonic copepod *Centropages typicus*. *Molecular Ecology*, **25**, 1566–1580.

Boehm, J.T., Waldman, J., Robinson, J.D. & Hickerson, M.J. (2015) Population genomics reveals seahorses (*Hippocampus erectus*) of the western mid-Atlantic coast to be residents rather than vagrants. *PLoS ONE*, **10**, e0116219.

Bryson, R.W., Savary, W.E., Zellmer, A.J., Bury, R.B. & McCormack, J.E. (2016) Genomic data reveal ancient microendemism in forest scorpions across the California Floristic Province. *Molecular Ecology*, **25**, 3731–3751.

Campagna, L., Gronau, I., Silveira, L.F., Siepel, A. & Lovette, I.J. (2015) Distinguishing noise from signal in patterns of genomic divergence in a highly polymorphic avian radiation. *Molecular Ecology*, **24**, 4238–4251.

Catchen, J.M., Amores, A., Hohenlohe, P.A., Cresko, W.A. & Postlethwait, J.H. (2011) *Stacks*: building and genotyping Loci *de novo* from short-read sequences. *G3: Genes, Genomes, Genetics*, **1**, 171–182.

Catchen, J., Bassham, S., Wilson, T., Currey, M., O'Brien, C., Yeates, Q. & Cresko, W.A. (2013b) The population structure and recent colonization history of Oregon threespine stickleback determined using restriction-site associated DNA-sequencing. *Molecular Ecology*, **22**, 2864–2883.

Catchen, J.M., Hohenlohe, P.A., Bassham, S., Amores, A. & Cresko, W.A. (2013a) *Stacks*: an analysis tool set for population genomics. *Molecular Ecology*, **22**, 3124–3140.

Combosch, D.J. & Vollmer, S.V. (2015) Trans-Pacific RAD-Seq population genomics confirms introgressive hybridization in Eastern Pacific Pocillopora corals. *Molecular Phylogenetics and Evolution*, **88**, 154–162.

Cristofari, R., Bertorelle, G., Ancel, A. *et al.* (2016b) Full circumpolar migration ensures evolutionary unity in the Emperor penguin. *Nature Communications*, **7**, 11842.

Cristofari, R., Liu, X., Bonadonna, F. *et al.* (2016a) Climate-driven range shifts in fragmented ecosystems. *bioRxiv*, https://doi.org/10.1101/090852.

Cristofari, R., Trucchi, E., Whittington, J.D., Vigetta, S., Gachot-Neveu, H., Stenseth, N.C., Le Maho, Y. & Le Bohec, C. (2015) Spatial heterogeneity as a genetic mixing mechanism in highly philopatric colonial seabirds. *PLoS ONE*, **10**, e0117981.

Davey, J.W., Cezard, T., Fuentes-Utrilla, P., Eland, C., Gharbi, K. & Blaxter, M.L. (2013) Special features of RAD Sequencing data: implications for genotyping. *Molecular Ecology*, **22**, 3151–3164.

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M. & Blaxter, M.L. (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, **12**, 499–510.

Díaz-Arce, N., Arrizabalaga, H., Murua, H., Irigoien, X. & Rodríguez-Ezpeleta, N. (2016) RAD-seq derived genome-wide nuclear markers

resolve the phylogeny of tunas. *Molecular Phylogenetics and Evolution*, **102**, 202–207.

Donnelly, R.K., Harper, G.L., Morgan, A.J., Pinto-Juma, G.A. & Bruford, M.W. (2014) Mitochondrial DNA and morphological variation in the sentinel earthworm species *Lumbricus rubellus*. *European Journal of Soil Biology*, **64**, 23–29.

Elo, K., Ivanoff, S., Vuorinen, J.A. & Piironen, J. (1997) Inheritance of RAPD markers and detection of interspecific hybridization with brown trout and Atlantic salmon. *Aquaculture*, **152**, 55–65.

Emerson, K.J., Merz, C.R., Catchen, J.M., Hohenlohe, P.A., Cresko, W.A., Bradshaw, W.E. & Holzapfel, C.M. (2010) Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences*, **107**, 16196–16200.

Epstein, B., Jones, M., Hamede, R. *et al.* (2016) Rapid evolutionary response to a transmissible cancer in Tasmanian devils. *Nature Communications*, **7**, 12684.

Etter, P.D., Preston, J.L., Bassham, S., Cresko, W.A. & Johnson, E.A. (2011) Local *de novo* assembly of RAD paired-end contigs using short sequencing reads. *PLoS ONE*, **6**, e18561.

Fountain, E.D., Pauli, J.N., Reid, B.N., Palsbøll, P.J. & Peery, M.Z. (2016) Finding the right coverage: the impact of coverage and sequence quality on SNP genotyping error rates. *Molecular Ecology Resources*, **16**, 966–978.

Fowler, B.L.S. & Buonaccorsi, V.P. (2016) Genomic characterization of sex-identification markers in *Sebastes carnatus* and *Sebastes chrysomelas* rockfishes. *Molecular Ecology*, **25**, 2165–2175.

Freer, J.J., Mable, B.K., Clucas, G., Rogers, A.D., Polito, M.J., Dunn, M., Naveen, R., Levy, H. & Hart, T. (2015) Limited genetic differentiation among chinstrap penguin (*Pygoscelis antarctica*) colonies in the Scotia Arc and Western Antarctic Peninsula. *Polar Biology*, **38**, 1493–1502.

Gaither, M.R., Bernal, M.A., Coleman, R.R., Bowen, B.W., Jones, S.A., Simison, W.B. & Rocha, L.A. (2015) Genomic signatures of geographic isolation and natural selection in coral reef fishes. *Molecular Ecology*, **24**, 1543–1557.

Gharbi, K., Gautier, A., Danzmann, R.G. *et al.* (2006) A linkage map for brown trout (*Salmo trutta*): chromosome homeologies and comparative genome organization with other salmonid fish. *Genetics*, **172**, 2405–2419.

Giska, I., Sechi, P. & Babik, W. (2015) Deeply divergent sympatric mitochondrial lineages of the earthworm *Lumbricus rubellus* are not reproductively isolated. *BMC Evolutionary Biology*, **15**, 217.

Graham, C.F., Glenn, T.C., McArthur, A.G. *et al.* (2015) Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). *Molecular Ecology Resources*, **15**, 1304–1315.

Hale, M.C., Thrower, F.P., Berntson, E.A., Miller, M.R. & Nichols, K.M. (2013) Evaluating adaptive divergence between migratory and nonmigratory ecotypes of a salmonid fish, *Oncorhynchus mykiss*. *G3: Genes| Genomes| Genetics*, **G3**, 1273–1285.

Ivy, J.A., Putnam, A.S., Navarro, A.Y., Gurr, J. & Ryder, O.A. (2016) Applying SNP-derived molecular coancestry estimates to captive breeding programs. *Journal of Heredity*, **107**, 403–412.

Jezkova, T., Riddle, B.R., Card, D.C., Schield, D.R., Eckstut, M.E. & Castoe, T.A. (2015) Genetic consequences of postglacial range expansion in two codistributed rodents (genus *Dipodomys*) depend on ecology and genetic locus. *Molecular Ecology*, **24**, 83–97.

Jones, J.C., Fan, S., Franchini, P., Schartl, M. & Meyer, A. (2013) The evolutionary history of *Xiphophorus* fish and their sexually selected sword: a genome-wide approach using restriction site-associated DNA sequencing. *Molecular Ecology*, **22**, 2986–3001.

Kautenburger, R. (2006) Genetic structure among earthworms (*Lumbricus terrestris* L.) from different sampling sites in western Germany based on random amplified polymorphic DNA. *Pedobiologia*, **50**, 257–266.

Keller, S., Levsen, N., Olson, M. & Tiffin, P. (2012) Local adaptation in the flowering-time gene network of balsam poplar, *Populus balsamifera* L. *Molecular Biology and Evolution*, **29**, 3143–3152.

King, R.A., Tibble, A.L. & Symondson, W.O.C. (2008) Opening a can of worms: unprecedented sympatric cryptic diversity within British lumbricid earthworms. *Molecular Ecology*, **17**, 4684–4698.

Kjeldsen, S.R., Zenger, K.R., Leigh, K., Ellis, W., Tobey, J., Phalen, D., Melzer, A., FitzGibbon, S. & Raadsma, H.W. (2016) Genome-wide SNP loci reveal novel insights into koala (*Phascolarctos cinereus*) population variability across its range. *Conservation Genetics*, **17**, 337–353.

Klarica, J., Kloss-Brandstätter, A., Traugott, M. & Juen, A. (2012) Comparing four mitochondrial genes in earthworms – Implications for identification, phylogenetics, and discovery of cryptic species. *Soil Biology and Biochemistry*, **45**, 23–30.

Ksepka, D.T., Bertelli, S. & Giannini, N.P. (2006) The phylogeny of the living and fossil Sphenisciformes (penguins). *Cladistics*, **22**, 412–441.

Laporte, M., Pavey, S.A., Rougeux, C. *et al.* (2016) RAD sequencing reveals within-generation polygenic selection in response to anthropogenic organic and metal contamination in North Atlantic Eels. *Molecular Ecology*, **25**, 219–237.

Leaniz, C.G. & Verspoor, E. (1989) Natural hybridization between Atlantic salmon, *Salmo salar*, and brown trout, *Salmo trutta*, in northern Spain. *Journal of Fish Biology*, **34**, 41–46.

Lemay, M.A. & Russello, M.A. (2015) Genetic evidence for ecological divergence in kokanee salmon. *Molecular Ecology*, **24**, 798–811.

Lescak, E.A., Bassham, S.L., Catchen, J., Gelmond, O., Sherbick, M.L., von Hippel, F.A. & Cresko, W.A. (2015) Evolution of stickleback in 50 years on earthquake-uplifted islands. *Proceedings of the National Academy of Sciences*, **112**, 201512020.

Longo, G. & Bernardi, G. (2015) The evolutionary history of the embiotocid surfperch radiation based on genome-wide RAD sequence data. *Molecular Phylogenetics and Evolution*, **88**, 55–63.

Lozier, J.D., Jackson, J.M., Dillon, M.E. & Strange, J.P. (2016) Population genomics of divergence among extreme and intermediate color forms in a polymorphic insect. *Ecology and Evolution*, **6**, 1075–1091.

Martin, C.H. & Feinstein, L.C. (2014) Novel trophic niches drive variable progress towards ecological speciation within an adaptive radiation of pupfishes. *Molecular Ecology*, **23**, 1846–1862.

Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T.H., Piñero, D. & Emerson, B.C. (2015) Restriction site-associated DNA sequencing, genotyping error estimation and *de novo* assembly optimization for population genetic inference. *Molecular Ecology Resources*, **15**, 28–41.

Nadeau, N.J., Whibley, A., Jones, R.T. *et al.* (2012) Genomic islands of divergence in hybridizing Heliconius butterflies identified by large-scale targeted sequencing. *Philosophical transactions of the Royal Society of London Series B, Biological Sciences*, **367**, 343–353.

Nims, B.D., Vargas, F.H., Merkel, J. & Parker, P.G. (2008) Low genetic diversity and lack of population structure in the endangered Galápagos penguin (*Spheniscus mendiculus*). *Conservation Genetics*, **9**, 1413–1420.

Palaiokostas, C., Bekaert, M., Taggart, J.B., Gharbi, K., McAndrew, B.J., Chatain, B., Penman, D.J. & Vandeputte, M. (2015) A new SNP-based vision of the genetics of sex determination in European sea bass (*Dicentrarchus labrax*). *Genetics Selection Evolution*, **47**, 68.

Pante, E., Abdelkrim, J., Viricel, A., Gey, D., France, S.C., Boisselier, M.C. & Samadi, S. (2015) Use of RAD sequencing for delimiting species. *Heredity*, **114**, 450–459.

Paris, J.R., King, R.A. & Stevens, J.R. (2015) Human mining activity across the ages determines the genetic structure of modern brown trout (*Salmo trutta* L.) populations. *Evolutionary Applications*, **8**, 573–585.

Pavey, S.A., Gaudin, J., Normandeau, E., Dionne, M., Castonguay, M., Audet, C. & Bernatchez, L. (2015) RAD sequencing highlights polygenic discrimination of habitat ecotypes in the panmictic American eel. *Current Biology*, **25**, 1666–1671.

Ravinet, M., Westram, A., Johannesson, K., Butlin, R., André, C. & Panova, M. (2016) Shared and nonshared genomic divergence in parallel ecotypes of *Littorina saxatilis* at a local scale. *Molecular Ecology*, **25**, 287–305.

Razkin, O., Sonet, G., Breugelmans, K., Madeira, M.J., Gómez-Moliner, B.J. & Backeljau, T. (2016) Species limits, interspecific hybridization and phylogeny in the cryptic land snail complex *Pyramidula*: the power of RADseq data. *Molecular Phylogenetics and Evolution*, **101**, 267–278.

Rodríguez-Ezpeleta, N., Bradbury, I.R., Mendibil, I., Álvarez, P., Cotano, U. & Irigoien, X. (2016) Population structure of Atlantic Mackerel inferred from RAD-seq derived SNP markers: effects of sequence clustering parameters and hierarchical SNP selection. *Molecular Ecology Resources*, **16**, 991–1001.

Roeder, A.D., Marshall, R.K., Mitchelson, A.J. *et al.* (2001) Gene flow on the ice: genetic differentiation among Adélie penguin colonies around Antarctica. *Molecular Ecology*, **10**, 1645–1656.

Rougemont, Q., Gagnaire, P.-A., Perrier, C., Genthon, C., Besnard, A.-L., Launey, S. & Evanno, G. (2016) Inferring the demographic history underlying parallel genomic divergence among pairs of parasitic and nonparasitic lamprey ecotypes. *Molecular Ecology*, **26**, 142–162.

Rowe, H.C., Renaut, S. & Guggisberg, A. (2011) RAD in the realm of next-generation sequencing technologies. *Molecular Ecology*, **20**, 3499–3502.

Saenz-Agudelo, P., Dibattista, J.D., Piatek, M.J., Gaither, M.R., Harrison, H.B., Nanninga, G.B. & Berumen, M.L. (2015) Seascape genetics along environmental gradients in the Arabian Peninsula: insights from ddRAD sequencing of anemonefishes. *Molecular Ecology*, **24**, 6241–6255.

Sechi, P. (2013) An evolutionary history of the peregrine epigeic earthworm *Lumbricus rubellus*. PhD thesis, Cardiff University, Cardiff, UK.

Shafer, A.B.A., Peart, C.R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C.W. & Wolf, J.B.W. (2016) Bioinformatic processing of RAD-seq data dramatically

impacts downstream population genetic inference. *Methods in Ecology and Evolution*, doi: 10.1111/2041-210X.12700.

Shepeleva, O.A., Kodolova, O.P., Zhukovskaya, E.A. & Striganova, B.R. (2008) Genetic diversity of populations of the earthworm *Lumbricus rubellus*. *Biology Bulletin*, **35**, 170–177.

Suyama, Y. & Matsuki, Y. (2015) MIG-seq: an effective PCR-based method for genome-wide single-nucleotide polymorphism genotyping using the next-generation sequencing platform. *Scientific Reports*, **5**, 16963.

Tariel, J., Longo, G.C. & Bernardi, G. (2016) Tempo and mode of speciation in Holacanthus angelfishes based on RADseq markers. *Molecular Phylogenetics and Evolution*, **98**, 84–88.

Trucchi, E., Frajman, B., Haverkamp, T., Schoenswetter, P. & Paun, O. (2016) Genomic and metagenomic analyses reveal parallel ecological divergence in *Helisperma pusillusm* (Carophyllaceae). *bioRxiv*, https://doi.org/10.1101/044354.

Wagner, C.E., Keller, I., Wittwer, S., Selz, O.M., Mwaiko, S., Greuter, L., Sivasundar, A. & Seehausen, O. (2013) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*, **22**, 787–798.

Wu, T.D. & Nacu, S. (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics*, **26**, 873–881.

Xu, P., Xu, S., Wu, X., Tao, Y., Wang, B., Wang, S., Qin, D., Lu, Z. & Li, G. (2014) Population genomic analyses from low-coverage RAD-Seq data: a case study on the non-model cucurbit bottle gourd. *The Plant Journal*, **77**, 430–442.

Received 21 January 2017; accepted 13 March 2017
Handling Editor: Susan Johnston

Supporting Information

Details of electronic Supporting Information are provided below.

Appendix S1. Additional Figures S1–S5 and Tables S1–S15.

Supporting Information

Lost in parameter space: A road map for Stacks

This supporting information contains:

Figure S1. K-mer distribution plots for *TRT*, *PGN*, and *ETW.*

Figure S2. Global SNP diversity statistics: nucleotide (Pi), gene and haplotype diversity calculated in *TRT*, *PGN*, and *ETW* with increasing values for the mismatch parameter (M) from M1-M6.

Figure S3. Plots of iterating values for the n parameter for *TRT*, *PGN* and *ETW* for n0-n10.

Figure S4. Plots comparing alignment to a reference genome (ref map) compared to creating loci *de novo* and integrating alignment positions (integrated) for the three datasets: *TRT*, *PGN* and *ETW*.

Figure S5. Plots showing the SNP frequency distributions for *TRT*, *PGN* and *ETW.*

Table S1. Number of reads for the 16 samples comprising each of the three datasets (*TRT*, *ETW* and *PGN*).

Table S2. Raw results for the mean and merged coverage with increasing values of m (the minimum number of raw reads required to form a stack).

Table S3. Raw results for the three datasets (*TRT*, *ETW* and *PGN*) for increasing values of m (the minimum number of raw reads required to form a stack).

Table S4. Combined analyses of iterating over values for m (m1-m6) showing the differences between M2 and M4.

Table S5. Combined analyses of iterating over values for M (M0-M6) showing the differences between m3 and m6.

Table S6. Raw results for the 16 samples from the three datasets (*TRT*, *PGN* and *ETW*) for increasing values of M (the maximum number of mismatches).

Table S7. Assembly metrics across 80% (r80) of the population for iterations of M0-M11 (the maximum number of mismatches) for *ETW.*

Table S8. Metrics collected for increasing values for M (the maximum number of mismatches) for each dataset (*TRT, PGN* and ETW) for the total number of polymorphic loci across 80% of the population (r80).

Table S9. Metrics for iterating over n (n0-n10) for each of the three datasets: *TRT, PGN* and *ETW.*

Table S10. Metrics resulting from testing n=M, n=M-1 and n=M+1 for the optimal value of M for each of the datasets.

Table S11. Raw results for the 16 samples for the *Stacks* default parameters (m3, M2, n1) and the optimal parameters for the three datasets: *TRT* (m3, M5, n4); *PGN* (m3, M2, n2); *ETW* (m3, M3, n3).

Table S12. Standard population genetics statistics: F_{IS}; H_E and F_{ST} calculated for each of the three datasets at 80% across the population.

Table S13. Global SNP diversity statistics: nucleotide (Pi); gene and haplotype diversity calculated for each of the three datasets (*TRT*, *PGN* and *ETW*) with increasing values for the mismatch parameter (M) from M1-M6.

Table S14. The number of consensus loci that aligned against the respective reference genomes using *ref map* and the *de novo - integrated* method.

Table S15. Raw data for the frequency of the number of SNPs found per catalog locus for different iterations of the n parameter (the number of mismatches allowed between stacks in the catalog).

Figure S1

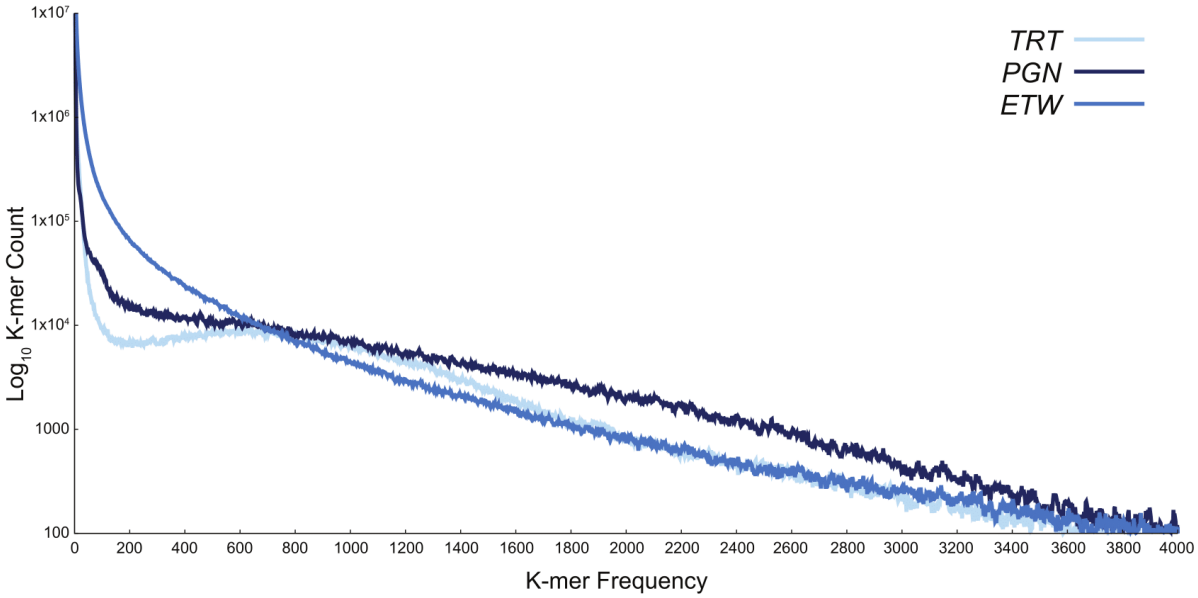


Figure S1. K-mer distribution plots for TRT, PGN, and ETW using a k-mer size of 15.

Figure S1. K-mer distribution plots for TRT, PGN, and ETW using a k-mer size of 15.

Figure S2

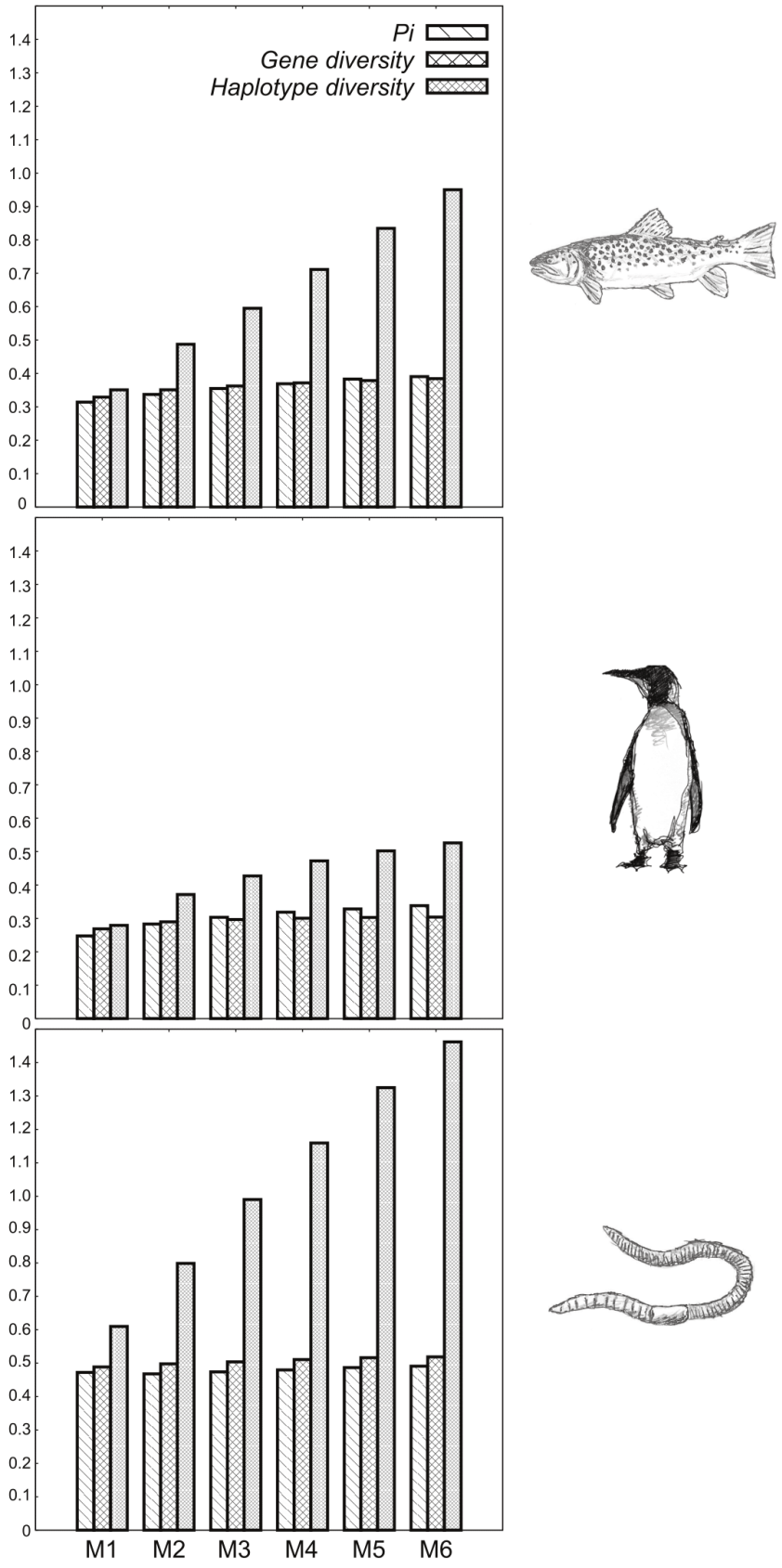


Figure S2. Global SNP diversity statistics: nucleotide (Pi), gene and haplotype diversity calculated in TRT, PGN, and ETW with increasing values for the mismatch parameter (M) from M1-M6.

Figure S3

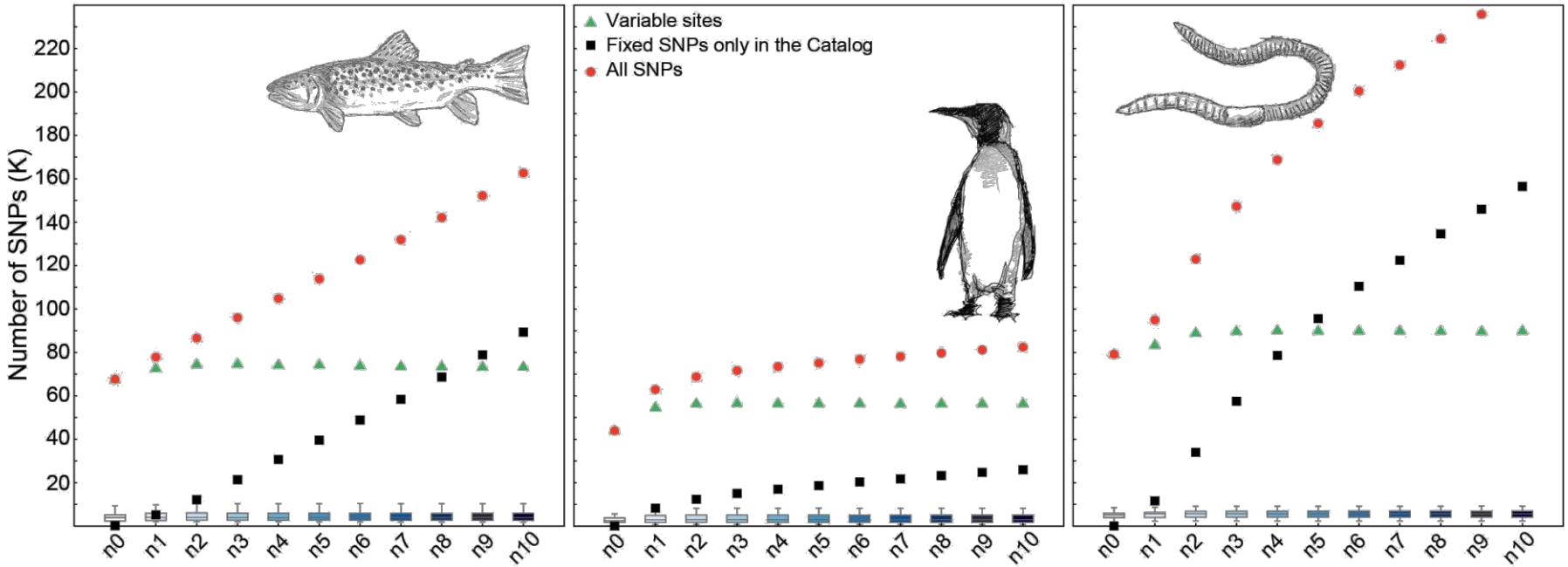


Figure S3. Plots of iterating values for the n parameter for TRT, PGN and ETW for n0-n10. Green triangles represent variable sites across the entire 16 samples in the dataset; red circles represent the number of SNPs across all samples and in the catalog and black squares represent the number of SNPs fixed only at the catalog level.

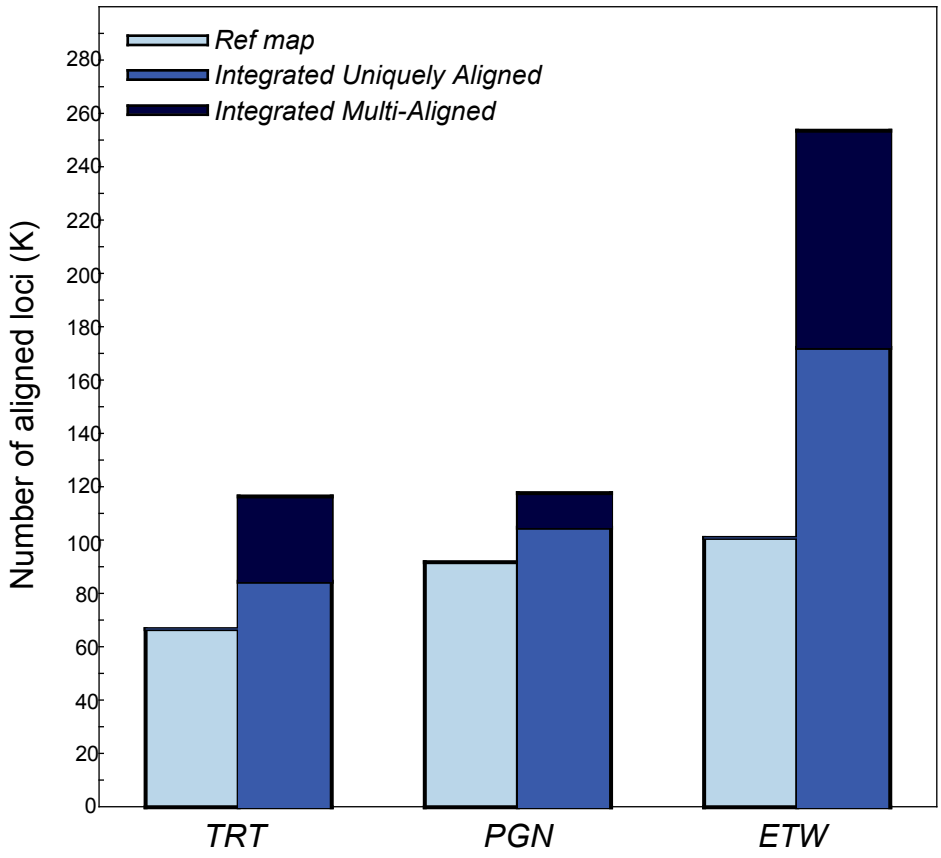


Figure S4. Plots comparing alignment to a reference genome (ref map) compared to creating loci *de novo* and integrating alignment positions (integrated) for the three datasets: TRT, PGN and ETW.

Figure S5

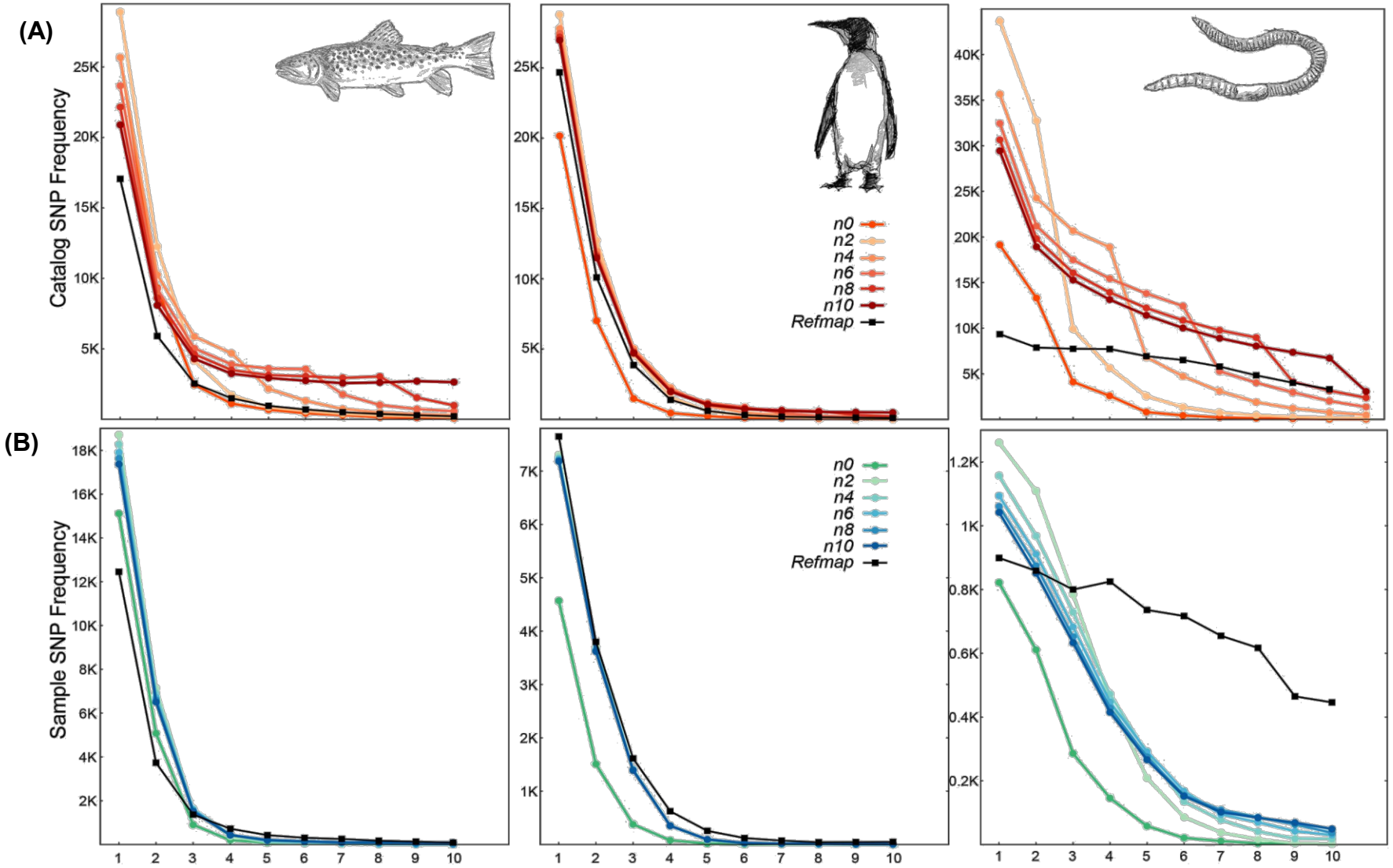


Figure S5. Plots showing the SNP frequency distributions for TRT, PGN and ETW for (A) catalog SNPs and (B) sample SNPs at *r80 loci*

Supporting Information Table S1. Number of reads for the 16 samples comprising each of the three datasets (*TRT*, *ETW* and *PGN*).

***TRT* brown trout dataset**		***PGN* king penguin dataset**		***ETW* red earthworm dataset**	
Sample	**Cleaned reads**	**Sample**	**Cleaned reads**	**Sample**	**Cleaned reads**
Sample 1	2,296,630	Sample 1	3,712,999	Sample 1	2,431,549
Sample 2	2,403,153	Sample 2	2,479,037	Sample 2	2,782,156
Sample 3	3,756,877	Sample 3	818,708	Sample 3	2,800,014
Sample 4	6,633,067	Sample 4	1,853,043	Sample 4	1,471,490
Sample 5	3,538,369	Sample 5	1,081,140	Sample 5	3,164,688
Sample 6	2,603,767	Sample 6	2,802,493	Sample 6	4,917,988
Sample 7	2,624,113	Sample 7	2,904,661	Sample 7	3,861,566
Sample 8	1,887,025	Sample 8	3,163,468	Sample 8	4,405,071
Sample 9	2,336,256	Sample 9	3,649,171	Sample 9	3,466,561
Sample 10	3,226,657	Sample 10	2,197,462	Sample 10	4,754,667
Sample 11	2,437,574	Sample 11	3,513,873	Sample 11	5,568,263
Sample 12	2,206,947	Sample 12	4,976,250	Sample 12	6,323,405
Sample 13	1,451,807	Sample 13	3,273,296	Sample 13	1,521,233
Sample 14	3,217,470	Sample 14	5,192,660	Sample 14	4,008,935
Sample 15	2,629,506	Sample 15	7,513,555	Sample 15	3,039,613
Sample 16	2,836,008	Sample 16	3,862,733	Sample 16	3,556,666
Total	46,085,226	**Total**	52,994,549	**Total**	58,073,865
Average	2,880,327	**Average**	3,312,159	**Average**	3,629,617
SE	290,590	**SE**	337,146	**SE**	409,701

Supporting Information Table S2. Raw results for the mean and merged coverage for each sample from the three datasets (*TRT*, *ETW* and *PGN*), with increasing values of m (the minimum number of raw reads required to form a stack). Other parameters are *Stacks* defaults (M2, n0).

***TRT* brown trout dataset**

	m1		m2		m3		m4		m5		m6	
	Mean	Mean merged	Mean	Mean merged	Mean	Mean merged	Mean	Mean merged	Mean	Mean merged	Mean	Mean merged
Sample 1	9	18.22	18	22.19	20	23.38	21	24.49	22	25.58	23	26.63
Sample 2	9	19.14	19	23.17	20	24.30	22	25.32	23	26.31	24	27.30
Sample 3	15	31.62	31	37.55	33	38.80	34	39.93	35	40.95	36	41.94
Sample 4	11	33.49	31	48.41	38	51.10	41	53.53	44	55.80	46	57.86
Sample 5	9	23.46	22	31.08	26	33.00	28	34.47	30	35.75	31	36.91
Sample 6	9	19.21	19	23.75	21	25.04	22	26.23	24	27.37	25	28.48
Sample 7	9	20.37	20	24.99	22	26.17	23	27.28	24	28.33	25	29.35
Sample 8	8	15.91	16	18.98	17	20.10	18	21.20	19	22.26	21	23.31
Sample 9	9	19.17	19	23.13	21	24.25	22	25.31	23	26.37	24	27.45
Sample 10	10	22.93	24	29.92	26	31.36	28	32.59	29	33.68	30	34.73
Sample 11	9	19.69	19	23.77	21	24.94	22	25.98	24	27.02	25	28.04
Sample 12	9	18.03	18	22.09	20	23.31	21	24.36	22	25.38	23	26.40
Sample 13	7	12.93	13	15.64	15	16.95	16	18.11	17	19.21	18	20.25
Sample 14	10	21.74	23	28.67	26	30.38	27	31.78	29	33.10	30	34.33
Sample 15	9	19.98	20	25.02	22	26.44	24	27.62	25	28.69	26	29.69
Sample 16	10	21.43	21	26.70	24	28.05	25	29.17	26	30.20	27	31.24
Average	9.50	21.08	20.81	26.57	23.25	27.97	24.63	29.21	26.00	30.38	27.13	31.49

***PGN* king penguin dataset**

	m1		m2		m3		m4		m5		m6	
	Mean	**Mean merged**	**Mean**	**Mean merged**	**Mean**	**Mean merged**	**Mean**	**Mean merged**	**Mean**	**Mean merged**	**Mean**	**Mean merged**
Sample 1	16	39.18	42	50.69	47	52.96	49	54.61	51	55.89	52	57.09
Sample 2	14	30.95	34	39.94	38	42.25	40	44.05	42	45.46	43	46.71
Sample 3	13	22.94	26	28.59	30	31.13	32	32.98	33	34.45	34	35.68
Sample 4	14	31.06	35	39.85	39	42.47	41	44.37	43	45.81	44	47.10
Sample 5	12	23.95	28	31.09	31	33.61	34	35.56	35	37.09	37	38.42
Sample 6	14	33.13	37	43.50	41	45.94	44	47.76	45	49.21	47	50.49
Sample 7	15	34.71	38	45.12	43	47.59	45	49.42	47	50.90	48	52.09
Sample 8	15	36.48	41	48.46	45	50.94	48	52.72	50	54.13	51	55.37
Sample 9	16	41.35	45	56.42	51	59.29	54	61.26	56	62.88	58	64.20
Sample 10	15	32.97	35	44.19	39	46.93	42	48.94	43	50.60	45	52.08
Sample 11	14	35.38	37	47.87	42	50.53	45	52.41	47	54.01	48	55.33
Sample 12	17	45.89	51	68.63	58	72.57	61	75.12	64	77.16	65	78.85
Sample 13	14	29.45	31	38.14	34	40.55	36	42.32	38	43.78	39	45.08
Sample 14	16	40.21	41	52.08	44	53.64	45	54.72	47	55.55	48	56.35
Sample 15	18	48.13	51	67.87	57	70.42	59	71.96	60	73.20	62	74.28
Sample 16	15	31.09	32	38.89	35	40.48	36	41.68	37	42.67	38	43.52
Average	14.88	34.80	37.75	46.33	42.13	48.83	44.44	50.62	46.13	52.05	47.44	53.29

***ETW* red earthworm dataset**

	m1		m2		m3		m4		m5		m6	
	Mean	**Mean merged**	**Mean**	**Mean merged**	**Mean**	**Mean merged**	**Mean**	**Mean merged**	**Mean**	**Mean merged**	**Mean**	**Mean merged**
Sample 1	5	6.14	12	12.90	16	17.40	20	20.97	22	23.79	24	26.03
Sample 2	5	6.76	13	14.09	18	18.93	21	22.91	25	26.25	27	28.93
Sample 3	5	6.41	12	13.26	17	17.80	20	21.82	24	25.21	26	27.98
Sample 4	4	5.55	10	11.17	14	14.69	16	17.15	18	19.05	19	20.64
Sample 5	5	6.72	13	14.21	18	19.11	22	23.32	25	26.96	28	29.97
Sample 6	6	8.31	16	17.80	22	23.15	26	27.82	30	32.21	34	36.07
Sample 7	6	9.76	17	19.11	23	24.94	27	29.40	31	32.97	34	36.00
Sample 8	5	7.27	13	14.97	18	19.69	22	23.59	25	27.03	28	30.07
Sample 9	5	7.54	14	15.49	19	20.46	23	24.66	26	28.23	29	31.27
Sample 10	6	9.55	18	19.72	24	25.82	29	30.81	33	35.05	36	38.69
Sample 11	6	9.41	18	19.85	24	25.75	29	30.85	33	35.43	37	39.48
Sample 12	7	10.32	20	22.06	26	28.37	31	33.75	36	38.75	40	43.31
Sample 13	4	5.71	10	11.33	14	14.80	16	17.28	18	19.20	20	20.78
Sample 14	5	7.41	15	16.03	20	21.24	24	25.95	28	30.17	32	33.78
Sample 15	5	6.79	13	14.17	18	18.99	22	23.23	25	26.79	28	29.73
Sample 16	5	7.20	14	15.29	19	20.29	23	24.86	27	28.87	30	32.23
Average	5.25	7.55	14.25	15.72	19.38	20.72	23.19	24.90	26.63	28.50	29.50	31.56

Supporting Information Table S3. Raw results for the 16 samples from the three datasets (*TRT*, *ETW* and *PGN*) for (a) the number of assembled loci; (b) the number of polymorphic loci and (c) the number of SNPs for increasing values of m (the minimum number of raw reads required to form a stack). Other parameters are *Stacks* defaults (M2, n0).

(a) Number of assembled loci *TRT* brown trout dataset

		m1	m2	m3	m4	m5	m6
ustacks	Sample 1	126,039	97,113	91,141	86,122	81,490	77,264
	Sample 2	125,539	97,502	91,994	87,464	83,306	79,315
	Sample 3	118,824	96,352	92,814	89,833	87,228	84,772
	Sample 4	198,031	129,020	120,793	114,440	109,039	104,484
	Sample 5	150,802	106,597	99,023	93,912	89,749	86,138
	Sample 6	135,570	103,019	96,681	91,378	86,642	82,259
	Sample 7	128,826	98,796	93,397	88,839	84,718	80,917
	Sample 8	118,595	93,195	86,839	81,283	76,232	71,532
	Sample 9	121,839	94,871	89,548	84,994	80,727	76,611
	Sample 10	140,695	101,455	95,873	91,659	88,055	84,739
	Sample 11	123,795	96,529	91,059	86,694	82,532	78,601
	Sample 12	122,429	93,713	87,805	83,204	78,960	74,952
	Sample 13	112,279	86,445	78,352	72,008	66,507	61,624
	Sample 14	147,980	105,418	98,438	93,347	88,891	84,965
	Sample 15	131,582	98,748	92,438	87,739	83,717	80,061
	Sample 16	132,345	99,903	94,134	89,823	86,040	82,420
	Average	127,433	98,125	92,626	88,289	84,218	80,489
cstacks	40% of individuals	79,890	85,463	84,224	82,764	81,006	79,102
	60% of individuals	59,773	74,443	73,424	71,512	69,278	66,830
	80% of individuals	29,440	61,714	60,357	57,388	54,071	50,708

(a) Number of assembled loci *PGN* king penguin dataset

		m1	m2	m3	m4	m5	m6
ustacks	Sample 1	94,779	70,255	66,890	64,669	63,007	61,508
	Sample 2	80,095	59,261	55,667	53,143	51,286	49,707
	Sample 3	35,686	27,388	24,942	23,379	22,234	21,330
	Sample 4	59,667	44,476	41,468	39,505	38,109	36,911
	Sample 5	45,143	33,036	30,297	28,443	27,100	25,999
	Sample 6	84,598	61,481	57,860	55,425	53,580	52,027
	Sample 7	83,686	61,620	58,080	55,720	53,904	52,503
	Sample 8	86,718	62,397	59,034	56,834	55,183	53,784
	Sample 9	88,254	62,040	58,718	56,652	55,035	53,771
	Sample 10	66,646	47,694	44,617	42,591	41,008	39,665
	Sample 11	99,303	70,198	66,064	63,407	61,291	59,603
	Sample 12	108,450	69,646	65,525	63,115	61,286	59,830
	Sample 13	109,509	81,063	75,713	72,156	69,398	67,048
	Sample 14	129,049	95,947	92,793	90,740	89,174	87,693
	Sample 15	156,113	106,696	102,434	100,029	98,151	96,536
	Sample 16	124,232	95,520	91,300	88,350	85,954	83,948
	Average	87,486	62,219	58,876	56,743	55,109	53,778
cstacks	40% of individuals	54,931	66,844	64,597	62,507	60,732	59,282
	60% of individuals	19,736	42,854	40,608	38,205	36,305	34,672
	80% of individuals	1,931	17,351	16,189	14,565	13,287	12,236

(a) Number of assembled loci *ETW* red earthworm dataset

		m1	m2	m3	m4	m5	m6
ustacks	Sample 1	389,570	158,352	111,420	88,961	76,116	67,840
	Sample 2	410,960	170,688	121,132	96,643	81,924	72,590
	Sample 3	430,219	178,978	126,816	99,339	83,244	73,036
	Sample 4	264,998	111,692	80,171	65,981	57,478	51,505
	Sample 5	470,782	192,568	136,641	108,020	90,625	79,548
	Sample 6	590,192	242,042	179,720	145,502	122,639	107,200
	Sample 7	394,871	180,502	133,928	111,020	97,166	87,542
	Sample 8	603,989	252,984	184,066	148,650	126,106	110,505
	Sample 9	458,880	194,422	141,169	113,480	96,568	85,201
	Sample 10	496,484	214,614	158,787	129,954	112,014	99,771
	Sample 11	585,919	247,464	184,909	150,717	128,486	113,204
	Sample 12	611,509	256,270	193,807	159,539	136,309	119,934
	Sample 13	266,104	114,097	82,593	67,961	59,178	53,137
	Sample 14	540,479	217,077	157,285	124,598	104,262	90,999
	Sample 15	446,844	185,615	132,098	104,065	87,605	76,991
	Sample 16	493,718	202,435	146,248	115,308	96,490	84,408
	Average	464,831	193,495	138,905	112,250	96,529	84,805
cstacks	40% of individuals	59,046	49,439	39,979	34,243	30,564	27,867
	60% of individuals	12,175	12,698	11,049	9,852	8,952	8,199
	80% of individuals	2,236	3,325	3,232	2,908	2,679	2,488

(b) Number of polymorphic loci *TRT* brown trout dataset

		m1	m2	m3	m4	m5	m6
ustacks	Sample 1	8,123	8,466	8,302	8,161	8,066	7,984
	Sample 2	8,438	9,019	8,872	8,743	8,627	8,425
	Sample 3	8,113	8,708	8,612	8,543	8,475	8,441
	Sample 4	15,792	15,128	14,872	14,610	14,434	14,225
	Sample 5	11,640	11,585	11,385	10,927	10,690	10,421
	Sample 6	8,660	9,020	8,984	8,812	8,699	8,550
	Sample 7	8,666	9,049	8,930	8,729	8,608	8,484
	Sample 8	7,192	7,651	7,525	7,380	7,263	7,125
	Sample 9	7,519	7,842	7,646	7,511	7,470	7,304
	Sample 10	8,952	9,192	9,068	8,933	8,858	8,742
	Sample 11	7,538	7,877	7,758	7,641	7,617	7,413
	Sample 12	7,492	7,820	7,672	7,522	7,363	7,212
	Sample 13	5,622	5,978	5,872	5,625	5,504	5,310
	Sample 14	8,975	9,229	9,079	8,860	8,719	8,616
	Sample 15	7,742	8,200	8,140	7,926	7,845	7,690
	Sample 16	8,225	8,607	8,512	8,343	8,193	8,003
	Average	8,174	8,658	8,562	8,443	8,334	8,214
cstacks	40% of individuals	16,575	29,098	29,423	29,166	28,822	28,340
	60% of individuals	11,932	25,795	26,214	25,752	25,172	24,434
	80% of individuals	4,788	20,998	21,549	20,663	19,633	18,508

(b) Number of polymorphic loci *PGN* king penguin dataset

		m1	m2	m3	m4	m5	m6
ustacks	Sample 1	3,985	4,263	4,149	4,070	4,025	3,991
	Sample 2	2,662	2,857	2,809	2,733	2,713	2,683
	Sample 3	622	699	671	661	646	634
	Sample 4	1,793	1,926	1,893	1,844	1,811	1,788
	Sample 5	904	961	923	890	880	854
	Sample 6	2,878	3,055	2,992	2,928	2,894	2,835
	Sample 7	2,960	3,163	3,110	3,046	2,987	2,964
	Sample 8	3,100	3,336	3,256	3,185	3,154	3,106
	Sample 9	3,948	4,222	4,190	4,107	4,059	4,026
	Sample 10	4,026	4,293	4,222	4,163	4,123	4,055
	Sample 11	5,511	5,797	5,691	5,599	5,511	5,477
	Sample 12	7,382	7,607	7,514	7,414	7,342	7,274
	Sample 13	6,596	7,003	6,864	6,735	6,665	6,572
	Sample 14	10,870	11,526	11,403	11,303	11,228	11,144
	Sample 15	12,552	13,290	13,141	13,010	12,943	12,820
	Sample 16	8,241	8,829	8,714	8,571	8,490	8,401
	Average	3,967	4,243	4,170	4,089	4,042	4,009
cstacks	40% of individuals	7,338	20,247	20,657	20,330	20,125	19,856
	60% of individuals	2,005	14,318	14,663	14,089	13,662	13,231
	80% of individuals	126	6,225	6,560	6,071	5,627	5,263

(b) Number of polymorphic loci *ETW* red earthworm dataset

		m1	m2	m3	m4	m5	m6
ustacks	Sample 1	4,433	5,162	4,747	4,150	3,876	3,558
	Sample 2	4,694	5,485	5,000	4,419	4,111	3,755
	Sample 3	5,020	5,916	5,410	4,680	4,322	3,964
	Sample 4	3,042	3,685	3,440	3,056	2,898	2,699
	Sample 5	5,390	6,196	5,640	4,888	4,530	4,122
	Sample 6	7,111	8,262	7,665	6,838	6,348	5,857
	Sample 7	5,278	6,055	5,696	5,108	4,839	4,539
	Sample 8	8,269	9,692	8,796	7,858	7,386	6,838
	Sample 9	5,485	6,402	5,824	5,136	4,811	4,448
	Sample 10	6,125	7,139	6,630	5,896	5,565	5,175
	Sample 11	7,315	8,493	7,769	7,034	6,599	6,051
	Sample 12	7,832	8,926	8,237	7,403	6,919	6,283
	Sample 13	3,216	3,932	3,639	3,235	3,036	2,848
	Sample 14	6,574	7,581	6,905	6,068	5,561	5,091
	Sample 15	5,266	6,101	5,606	4,863	4,541	4,153
	Sample 16	5,828	6,685	6,155	5,423	4,987	4,543
	Average	5,438	6,299	5,760	5,122	4,825	4,494
cstacks	40% of individuals	4,448	10,666	10,670	9,546	8,855	8,220
	60% of individuals	1,618	4,849	4,975	4,605	4,307	3,976
	80% of individuals	383	1,771	1,972	1,803	1,688	1,587

(c) Number of SNPs *TRT* brown trout dataset

		m1	m2	m3	m4	m5	m6
ustacks	Sample 1	14,175	13,929	13,643	13,573	13,549	13,673
	Sample 2	14,408	14,774	14,258	14,396	14,322	14,025
	Sample 3	14,206	14,639	14,419	14,251	14,192	14,376
	Sample 4	25,965	22,902	22,281	21,917	21,797	21,596
	Sample 5	19,441	17,927	17,464	16,725	16,552	16,176
	Sample 6	15,224	14,758	14,639	14,472	14,463	14,450
	Sample 7	15,012	14,842	14,674	14,341	14,280	14,338
	Sample 8	12,295	12,617	12,350	12,270	12,158	12,247
	Sample 9	13,112	12,901	12,473	12,456	12,594	12,425
	Sample 10	16,233	15,114	14,939	14,773	14,842	14,855
	Sample 11	13,337	13,110	12,882	12,780	13,052	12,793
	Sample 12	13,443	13,252	13,040	12,871	12,565	12,635
	Sample 13	10,088	10,115	9,834	9,697	9,656	9,487
	Sample 14	16,099	15,214	14,900	14,621	14,528	14,619
	Sample 15	13,775	13,649	13,711	13,374	13,292	13,357
	Sample 16	14,658	14,212	14,136	13,897	13,755	13,414
	Average	14,307	14,426	14,197	14,074	13,974	13,849
cstacks	40% of individuals	21,139	44,706	46,499	47,037	46,999	46,828
	60% of individuals	14,457	37,517	39,067	38,941	38,707	38,002
	80% of individuals	5,513	28,581	30,247	29,203	28,016	26,777

(c) Number of SNPs *PGN* king penguin dataset

		m1	m2	m3	m4	m5	m6
ustacks	Sample 1	5,269	5,778	5,595	5,468	5,422	5,374
	Sample 2	3,621	3,970	3,916	3,810	3,774	3,736
	Sample 3	826	916	889	889	848	848
	Sample 4	2,374	2,593	2,581	2,508	2,453	2,419
	Sample 5	1,303	1,346	1,296	1,220	1,228	1,190
	Sample 6	3,985	4,298	4,188	4,146	4,067	4,033
	Sample 7	4,063	4,450	4,379	4,259	4,075	4,080
	Sample 8	4,167	4,615	4,493	4,404	4,357	4,261
	Sample 9	5,007	5,462	5,473	5,322	5,251	5,180
	Sample 10	5,152	5,632	5,526	5,466	5,418	5,327
	Sample 11	7,027	7,519	7,390	7,219	7,150	7,107
	Sample 12	9,478	9,986	9,818	9,672	9,584	9,487
	Sample 13	8,706	9,347	9,069	8,942	8,937	8,706
	Sample 14	13,747	14,921	14,827	14,681	14,550	14,353
	Sample 15	16,210	17,549	17,283	17,111	17,031	16,710
	Sample 16	10,880	11,970	11,803	11,657	11,503	11,358
	Average	5,080	5,547	5,500	5,394	5,335	5,254
cstacks	40% of individuals	8,433	27,993	29,062	28,723	28,690	28,311
	60% of individuals	2,283	19,607	20,477	19,697	19,267	18,777
	80% of individuals	142	8,539	9,168	8,545	7,994	7,528

(c) Number of SNPs *ETW* red earthworm dataset

		m1	m2	m3	m4	m5	m6
ustacks	Sample 1	6,851	8,105	7,668	6,957	6,504	6,005
	Sample 2	7,169	8,540	8,080	7,358	7,056	6,430
	Sample 3	7,578	9,066	8,780	7,743	7,261	6,687
	Sample 4	4,658	5,630	5,489	5,016	4,885	4,601
	Sample 5	8,401	9,714	9,210	8,270	7,852	7,131
	Sample 6	10,887	12,965	12,564	11,554	11,022	10,260
	Sample 7	8,287	9,488	9,225	8,418	8,091	7,681
	Sample 8	12,873	14,995	14,197	13,075	12,619	11,838
	Sample 9	8,429	10,048	9,612	8,716	8,357	7,754
	Sample 10	9,534	11,268	10,957	9,990	9,567	8,959
	Sample 11	11,502	13,422	12,742	11,989	11,370	10,587
	Sample 12	12,282	14,038	13,581	12,527	11,967	10,849
	Sample 13	4,687	6,036	5,778	5,250	4,995	4,714
	Sample 14	10,170	11,756	11,251	10,130	9,502	8,867
	Sample 15	7,830	9,375	9,086	8,067	7,647	7,124
	Sample 16	8,997	10,441	10,053	9,069	8,475	7,821
	Average	8,415	9,881	9,419	8,567	8,224	7,718
cstacks	40% of individuals	6,789	21,127	22,331	20,991	20,020	18,821
	60% of individuals	2,435	9,797	10,732	10,370	9,913	9,207
	80% of individuals	560	3,502	4,098	3,825	3,677	3,493

Supporting Information Table S4. Combined analyses of iterating over values for m (m1-m6) showing the differences between M2 and M4 for (a) Number of assembled loci (b) Number of polymorphic loci and (c) Number of SNPs for the three datasets: *TRT, PGN* and *ETW*.

(a) Number of assembled loci

ustacks							**cstacks (80% of individuals)**						
M2	**m1**	**m2**	**m3**	**m4**	**m5**	**m6**	**M2**	**m1**	**m2**	**m3**	**m4**	**m5**	**m6**
TRT	127,433	98,125	92,626	88,289	84,218	80,489	*TRT*	29,440	61,714	60,357	57,388	54,071	50,708
PGN	464,831	193,495	138,905	112,250	96,529	84,805	*PGN*	2,236	3,325	3,232	2,908	2,679	2,488
ETW	87,486	62,219	58,876	56,743	55,109	53,778	*ETW*	1,931	17,351	16,189	14,565	13,287	12,236
M4	**m1**	**m2**	**m3**	**m4**	**m5**	**m6**	**M4**	**m1**	**m2**	**m3**	**m4**	**m5**	**m6**
TRT	126,592	96,552	91,019	86,775	82,785	79,111	*TRT*	28,369	58,213	57,373	54,685	51,589	48,342
PGN	458,294	190,957	136,990	110,680	95,140	83,580	*PGN*	2,086	3,312	3,220	2,901	2,690	2,488
ETW	86,578	62,077	58,716	56,594	54,971	53,647	*ETW*	1,947	17,217	16,042	14,461	13,186	12,136

(b) Number of polymorphic loci

ustacks							**cstacks (80% of individuals)**						
M2	**m1**	**m2**	**m3**	**m4**	**m5**	**m6**	**M2**	**m1**	**m2**	**m3**	**m4**	**m5**	**m6**
TRT	8,174	8,658	8,562	8,443	8,334	8,214	*TRT*	4,788	20,998	21,549	20,663	19,633	18,508
PGN	5,438	6,299	5,760	5,122	4,825	4,494	*PGN*	383	1,771	1,972	1,803	1,688	1,587
ETW	3,967	4,243	4,170	4,089	4,042	4,009	*ETW*	126	6,225	6,560	6,071	5,627	5,263
M4	**m1**	**m2**	**m3**	**m4**	**m5**	**m6**	**M4**	**m1**	**m2**	**m3**	**m4**	**m5**	**m6**
TRT	9,148	9,986	9,971	9,882	9,769	9,657	*TRT*	4,668	20,794	21,662	20,885	19,983	18,828
PGN	6,411	8,539	7,742	6,835	6,354	5,882	*PGN*	385	1,908	2,083	1,901	1,795	1,679
ETW	4,075	4,398	4,313	4,230	4,172	4,145	*ETW*	130	6,189	6,508	6,037	5,584	5,214

(c) Number of SNPs													
ustacks							**cstacks (80% of individuals**						
M2	**m1**	**m2**	**m3**	**m4**	**m5**	**m6**	**M2**	**m1**	**m2**	**m3**	**m4**	**m5**	**m6**
TRT	14,307	14,426	14,197	14,074	13,974	13,849	*TRT*	5,513	28,581	30,247	29,203	28,016	26,777
PGN	8,415	9,881	9,419	8,567	8,224	7,718	*PGN*	560	3,502	4,098	3,825	3,677	3,493
ETW	5,080	5,547	5,500	5,394	5,335	5,254	*ETW*	142	8,539	9,168	8,545	7,994	7,528
M4	**m1**	**m2**	**m3**	**m4**	**m5**	**m6**	**M4**	**m1**	**m2**	**m3**	**m4**	**m5**	**m6**
TRT	22,806	23,131	23,111	23,156	22,816	22,603	*TRT*	6,113	35,837	39,352	38,706	37,606	35,852
PGN	13,425	18,791	17,678	15,876	14,845	13,842	*PGN*	650	4,843	5,541	5,161	5,019	4,753
ETW	5,939	6,594	6,534	6,474	6,313	6,298	*ETW*	151	8,696	9,337	8,735	8,140	7,650

Supporting Information Table S5. Combined analyses of iterating over values for M (M0-M6) showing the differences between m3 and m6 for (a) Number of assembled loci (b) Number of polymorphic loci and (c) Number of SNPs for the three datasets: *TRT, PGN* and *ETW*.

(a) Number of assembled loci

ustacks								**cstacks (80% of individuals)**							
m3	**M0**	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**m3**	**M0**	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**
TRT	107,246	94,634	92,626	91,750	91,019	90,504	90,036	*TRT*	67,439	63,157	60,343	58,545	57,370	56,399	55,552
PGN	146,984	140,746	138,905	137,842	136,990	136,381	135,903	*PGN*	3,615	3,094	3,227	3,216	3,220	3,231	3,244
ETW	65,958	59,621	58,876	58,751	58,716	58,695	58,669	*ETW*	17,600	16,261	16,193	16,090	16,045	16,028	16,006
m6	**M0**	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**m6**	**M0**	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**
TRT	90,586	82,164	80,489	79,748	79,111	78,608	78,181	*TRT*	54,678	52,891	50,714	49,257	48,343	47,610	46,953
PGN	89,219	85,961	84,805	84,080	83,580	83,255	82,971	*PGN*	2,420	2,371	2,487	2,473	2,485	2,499	2,510
ETW	58,603	54,405	53,778	53,668	53,647	53,626	53,603	*ETW*	12,651	12,228	12,240	12,164	12,135	12,127	12,117

(b) Number of polymorphic loci

ustacks								**cstacks (80% of individuals)**							
m3	**M0**	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**m3**	**M0**	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**
TRT	414	6,442	8,582	9,354	9,969	10,372	10,758	*TRT*	482	20,416	21,528	21,579	21,645	21,682	21,673
PGN	960	3,936	5,741	6,907	7,734	8,333	8,854	*PGN*	120	1,598	1,964	2,033	2,082	2,122	2,156
ETW	226	3,555	4,166	4,268	4,312	4,344	4,351	*ETW*	66	6,396	6,562	6,533	6,508	6,508	6,513
m6	**M0**	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**m6**	**M0**	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**
TRT	1,395	6,231	8,226	9,007	9,648	10,112	10,517	*TRT*	2,675	17,435	18,496	18,634	18,847	18,970	19,141
PGN	1,083	3,196	4,507	5,263	5,887	6,297	6,652	*PGN*	186	1,312	1,586	1,629	1,678	1,711	1,737
ETW	352	3,411	4,007	4,099	4,148	4,167	4,191	*ETW*	198	5,093	5,266	5,228	5,212	5,216	5,224

(c) Number of SNPs

ustacks								**cstacks (80% of individuals)**							
m3	**M0**	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**m3**	**M0**	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**
TRT	456	7,058	14,234	18,780	23,145	27,252	31,587	*TRT*	515	24,111	30,194	34,501	39,223	43,995	49,499
PGN	1,159	4,890	9,420	13,694	17,647	21,420	24,826	*PGN*	130	2,426	4,076	4,830	5,520	6,116	6,618
ETW	247	3,770	5,503	6,110	6,547	6,918	7,181	*ETW*	69	8,258	9,161	9,310	9,316	9,452	9,566
m6	**M0**	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**m6**	**M0**	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**
TRT	1,652	7,063	13,876	18,296	22,561	26,895	31,069	*TRT*	2,960	21,162	26,762	31,098	35,961	40,523	46,182
PGN	1,419	4,287	7,722	10,729	13,769	16,406	18,985	*PGN*	261	2,120	3,486	4,114	4,729	5,267	5,753
ETW	401	3,627	5,252	5,859	6,245	6,583	6,855	*ETW*	209	6,696	7,567	7,620	7,631	7,733	7,884

Supporting Information Table S6. Raw results for the 16 samples from the three datasets (*TRT*, *PGN* and *ETW*) for (a) the number of assembled loci; (b) the number of polymorphic loci and (c) the number of SNPs for increasing values of M (the maximum number of mismatches). Other parameters are *Stacks* defaults (m3, n0).

(a) Number of assembled loci *TRT* brown trout dataset

		M0	M1	M2	M3	M4	M5	M6
ustacks	Sample 1	105,114	93,065	91,141	90,277	89,625	89,066	88,647
	Sample 2	106,634	94,018	91,994	91,128	90,409	89,897	89,447
	Sample 3	107,858	94,886	92,814	91,931	91,212	90,723	90,304
	Sample 4	160,651	123,075	120,793	119,847	119,194	118,613	118,155
	Sample 5	123,183	100,943	99,023	98,204	97,652	97,207	96,786
	Sample 6	112,002	98,803	96,681	95,832	95,084	94,498	93,995
	Sample 7	108,708	95,533	93,397	92,528	91,808	91,266	90,815
	Sample 8	99,150	88,618	86,839	86,039	85,450	84,932	84,524
	Sample 9	102,426	91,448	89,548	88,723	88,056	87,552	87,096
	Sample 10	112,265	97,991	95,873	94,920	94,207	93,717	93,243
	Sample 11	104,294	92,990	91,059	90,203	89,549	88,979	88,520
	Sample 12	100,784	89,653	87,805	86,953	86,322	85,820	85,382
	Sample 13	87,421	79,858	78,352	77,671	77,087	76,638	76,229
	Sample 14	114,538	100,585	98,438	97,548	96,826	96,269	95,807
	Sample 15	106,597	94,381	92,438	91,569	90,826	90,284	89,768
	Sample 16	109,072	96,190	94,134	93,241	92,519	92,001	91,565
	Average	107,246	94,634	92,626	91,750	91,019	90,504	90,036
cstacks	40% of individuals	99,465	87,147	84,266	83,128	82,301	81,611	81,135
	60% of individuals	84,244	76,247	73,432	72,159	71,224	70,351	69,645
	80% of individuals	67,439	63,157	60,343	58,545	57,370	56,399	55,552

(a) Number of assembled loci *PGN* king penguin dataset

		M0	M1	M2	M3	M4	M5	M6
ustacks	Sample 1	74,683	67,693	66,890	66,758	66,725	66,718	66,713
	Sample 2	60,629	56,215	55,667	55,574	55,554	55,535	55,513
	Sample 3	26,022	25,061	24,942	24,909	24,879	24,869	24,848
	Sample 4	44,910	41,862	41,468	41,389	41,365	41,347	41,331
	Sample 5	32,011	30,502	30,297	30,251	30,237	30,219	30,206
	Sample 6	63,668	58,486	57,860	57,767	57,766	57,728	57,697
	Sample 7	63,858	58,710	58,080	57,978	57,968	57,962	57,932
	Sample 8	65,199	59,707	59,034	58,922	58,916	58,906	58,877
	Sample 9	66,716	59,535	58,718	58,579	58,516	58,484	58,460
	Sample 10	52,444	45,382	44,617	44,499	44,441	44,410	44,383
	Sample 11	77,581	67,047	66,064	65,874	65,818	65,769	65,732
	Sample 12	80,230	66,783	65,525	65,309	65,227	65,160	65,086
	Sample 13	87,832	76,900	75,713	75,521	75,460	75,422	75,378
	Sample 14	111,922	94,533	92,793	92,567	92,491	92,452	92,400
	Sample 15	125,221	104,421	102,434	102,109	102,029	101,985	101,942
	Sample 16	105,386	92,804	91,300	91,032	90,952	90,922	90,857
	Average	65,958	59,621	58,876	58,751	58,716	58,695	58,669
cstacks	40% of individuals	72,875	65,581	64,587	64,398	64,297	64,231	64,200
	60% of individuals	45,087	41,092	40,613	40,411	40,306	40,263	40,220
	80% of individuals	17,600	16,261	16,193	16,090	16,045	16,028	16,006

(a) Number of assembled loci *ETW* red earthworm dataset

		M0	M1	M2	M3	M4	M5	M6
ustacks	Sample 1	117,739	112,920	111,420	110,533	109,936	109,431	109,068
	Sample 2	128,046	122,723	121,132	120,170	119,517	119,026	118,596
	Sample 3	134,296	128,463	126,816	125,856	125,159	124,654	124,251
	Sample 4	84,442	81,266	80,171	79,564	79,148	78,818	78,555
	Sample 5	144,822	138,420	136,641	135,669	134,878	134,238	133,783
	Sample 6	191,122	182,294	179,720	178,241	177,111	176,307	175,654
	Sample 7	142,574	135,923	133,928	132,887	132,143	131,584	131,087
	Sample 8	197,806	187,277	184,066	182,150	180,782	179,657	178,835
	Sample 9	149,145	143,071	141,169	140,014	139,101	138,523	138,023
	Sample 10	168,939	161,138	158,787	157,407	156,490	155,788	155,203
	Sample 11	197,106	187,555	184,909	183,372	182,213	181,352	180,658
	Sample 12	206,970	196,646	193,807	192,198	191,017	190,109	189,370
	Sample 13	87,150	83,762	82,593	81,966	81,476	81,138	80,886
	Sample 14	167,399	159,451	157,285	155,955	155,068	154,361	153,737
	Sample 15	139,870	133,820	132,098	130,979	130,225	129,641	129,201
	Sample 16	155,156	148,119	146,248	145,051	144,224	143,601	143,117
	Average	146,984	140,746	138,905	137,842	136,990	136,381	135,903
cstacks	40% of individuals	44,016	40,591	39,994	39,678	39,498	39,392	39,311
	60% of individuals	12,159	10,976	11,064	11,077	11,137	11,190	11,175
	80% of individuals	3,615	3,094	3,227	3,216	3,220	3,231	3,244

(b) Number of polymorphic loci *TRT* brown trout dataset

		M0	M1	M2	M3	M4	M5	M6
ustacks	Sample 1	416	6,300	8,324	9,117	9,730	10,184	10,559
	Sample 2	420	6,709	8,876	9,600	10,220	10,664	11,019
	Sample 3	300	6,477	8,653	9,385	9,978	10,355	10,739
	Sample 4	730	12,359	14,871	15,794	16,460	16,994	17,481
	Sample 5	697	9,268	11,338	12,262	12,794	13,267	13,684
	Sample 6	423	6,863	8,955	9,756	10,417	10,855	11,303
	Sample 7	447	6,691	8,926	9,755	10,323	10,822	11,242
	Sample 8	400	5,587	7,553	8,235	8,728	9,212	9,549
	Sample 9	419	5,806	7,681	8,470	8,995	9,454	9,880
	Sample 10	412	6,778	9,082	9,891	10,531	10,965	11,317
	Sample 11	355	5,799	7,764	8,572	9,152	9,620	9,999
	Sample 12	387	5,669	7,696	8,418	9,050	9,490	9,867
	Sample 13	405	4,237	5,913	6,609	7,210	7,700	8,148
	Sample 14	467	6,875	9,032	9,867	10,511	10,965	11,438
	Sample 15	395	5,965	8,155	9,012	9,696	10,182	10,711
	Sample 16	407	6,406	8,510	9,323	9,960	10,389	10,776
	Average	414	6,442	8,582	9,354	9,969	10,372	10,758
cstacks	40% of individuals	1,077	27,075	29,458	30,243	30,721	31,197	31,552
	60% of individuals	782	24,506	26,211	26,719	27,020	27,239	27,441
	80% of individuals	482	20,416	21,528	21,579	21,645	21,682	21,673

(b) Number of polymorphic loci *PGN* king penguin dataset

		M0	M1	M2	M3	M4	M5	M6
ustacks	Sample 1	247	3,559	4,145	4,255	4,274	4,308	4,291
	Sample 2	226	2,355	2,801	2,881	2,909	2,912	2,936
	Sample 3	85	566	669	704	720	731	746
	Sample 4	171	1,584	1,888	1,971	1,994	2,017	2,031
	Sample 5	115	778	925	965	978	993	1,004
	Sample 6	213	2,522	2,984	3,080	3,095	3,124	3,141
	Sample 7	194	2,631	3,100	3,211	3,234	3,232	3,275
	Sample 8	204	2,763	3,251	3,361	3,388	3,401	3,405
	Sample 9	225	3,550	4,186	4,281	4,350	4,379	4,411
	Sample 10	224	3,681	4,230	4,339	4,395	4,435	4,462
	Sample 11	278	4,959	5,672	5,848	5,927	5,952	5,991
	Sample 12	332	6,614	7,517	7,738	7,848	7,896	7,942
	Sample 13	371	5,991	6,875	7,084	7,153	7,193	7,221
	Sample 14	328	10,123	11,406	11,626	11,733	11,750	11,780
	Sample 15	468	11,582	13,141	13,438	13,550	13,575	13,635
	Sample 16	334	7,513	8,716	8,937	9,033	9,063	9,095
	Average	226	3,555	4,166	4,268	4,312	4,344	4,351
cstacks	40% of individuals	433	19,860	20,649	20,682	20,647	20,627	20,659
	60% of individuals	214	14,298	14,656	14,631	14,580	14,576	14,570
	80% of individuals	66	6,396	6,562	6,533	6,508	6,508	6,513

(b) Number of polymorphic loci *ETW* red earthworm dataset

		M0	M1	M2	M3	M4	M5	M6
ustacks	Sample 1	723	3,304	4,763	5,652	6,274	6,787	7,174
	Sample 2	848	3,495	4,996	6,005	6,704	7,185	7,666
	Sample 3	886	3,785	5,399	6,430	7,140	7,700	8,153
	Sample 4	524	2,370	3,436	4,076	4,501	4,877	5,174
	Sample 5	955	3,915	5,624	6,632	7,484	8,126	8,628
	Sample 6	1,120	5,259	7,687	9,134	10,330	11,165	11,766
	Sample 7	774	3,866	5,675	6,683	7,484	8,044	8,505
	Sample 8	1,361	6,099	8,784	10,456	11,757	12,796	13,538
	Sample 9	901	3,956	5,807	7,025	7,943	8,529	9,051
	Sample 10	990	4,482	6,625	8,016	8,953	9,735	10,314
	Sample 11	1,097	5,302	7,778	9,281	10,469	11,308	12,006
	Sample 12	1,188	5,592	8,230	9,817	11,054	11,948	12,668
	Sample 13	566	2,513	3,635	4,293	4,766	5,153	5,434
	Sample 14	1,206	4,832	6,900	8,319	9,292	10,044	10,696
	Sample 15	987	3,902	5,606	6,788	7,525	8,137	8,656
	Sample 16	965	4,292	6,160	7,409	8,310	8,941	9,478
	Average	960	3,936	5,741	6,907	7,734	8,333	8,854
cstacks	40% of individuals	1,350	8,307	10,672	11,965	12,833	13,452	13,922
	60% of individuals	479	4,054	4,979	5,395	5,686	5,858	5,967
	80% of individuals	120	1,598	1,964	2,033	2,082	2,122	2,156

(c) Number of SNPs *TRT* brown trout dataset

		M0	M1	M2	M3	M4	M5	M6
ustacks	Sample 1	462	6,829	13,583	18,064	22,428	26,601	30,785
	Sample 2	454	7,279	14,347	18,793	23,275	27,373	31,493
	Sample 3	329	7,071	14,536	18,767	23,256	27,131	31,214
	Sample 4	777	13,542	22,266	28,007	33,614	39,362	45,192
	Sample 5	758	10,184	17,331	22,268	26,791	31,459	36,240
	Sample 6	456	7,472	14,525	19,184	23,991	28,213	32,775
	Sample 7	498	7,313	14,558	19,401	23,699	28,025	32,544
	Sample 8	429	6,070	12,473	16,336	19,922	24,364	28,049
	Sample 9	457	6,344	12,599	16,755	20,547	25,054	29,021
	Sample 10	460	7,397	14,976	19,769	24,146	28,442	32,724
	Sample 11	392	6,358	12,938	17,103	21,450	25,342	29,518
	Sample 12	428	6,236	13,076	16,943	21,081	25,127	28,533
	Sample 13	455	4,650	10,175	13,639	17,168	20,999	24,783
	Sample 14	504	7,596	14,682	19,653	24,188	28,896	33,580
	Sample 15	454	6,556	13,727	18,276	22,850	27,126	31,688
	Sample 16	448	7,045	14,120	18,960	23,034	27,568	31,681
	Average	456	7,058	14,234	18,780	23,145	27,252	31,587
cstacks	40% of individuals	1,180	32,804	46,773	56,715	66,712	77,023	88,305
	60% of individuals	845	29,376	39,155	47,047	54,801	62,373	71,078
	80% of individuals	515	24,111	30,194	34,501	39,223	43,995	49,499

(c) Number of SNPs *PGN* king penguin dataset

		M0	M1	M2	M3	M4	M5	M6
ustacks	Sample 1	270	3,783	5,616	6,506	6,935	7,436	7,454
	Sample 2	246	2,535	3,856	4,462	4,920	5,106	5,346
	Sample 3	90	616	869	1,046	1,182	1,286	1,389
	Sample 4	181	1,707	2,573	2,939	3,246	3,378	3,547
	Sample 5	118	846	1,291	1,578	1,676	1,831	1,943
	Sample 6	232	2,711	4,166	4,964	5,313	5,451	5,792
	Sample 7	211	2,833	4,336	4,992	5,387	5,614	6,027
	Sample 8	222	2,938	4,506	5,230	5,683	6,028	6,188
	Sample 9	242	3,756	5,450	6,043	6,600	6,840	7,140
	Sample 10	248	3,992	5,555	6,177	6,493	6,995	7,221
	Sample 11	299	5,346	7,346	8,184	8,933	9,238	9,694
	Sample 12	365	7,260	9,834	10,999	11,902	12,385	12,865
	Sample 13	399	6,481	9,071	10,503	11,103	11,564	12,003
	Sample 14	354	10,829	14,811	16,449	17,832	18,365	18,828
	Sample 15	510	12,460	17,255	19,227	20,756	21,358	22,073
	Sample 16	362	8,079	11,762	13,345	14,517	15,118	15,677
	Average	247	3,770	5,503	6,110	6,547	6,918	7,181
cstacks	40% of individuals	471	24,661	28,944	30,620	31,636	32,229	33,051
	60% of individuals	239	18,084	20,369	20,912	21,266	21,584	21,958
	80% of individuals	69	8,258	9,161	9,310	9,316	9,452	9,566

(c) Number of SNPs *ETW* red earthworm dataset

		M0	M1	M2	M3	M4	M5	M6
ustacks	Sample 1	872	4,080	7,737	11,040	14,192	17,156	20,292
	Sample 2	1,018	4,305	8,113	11,868	15,242	18,325	21,470
	Sample 3	1,062	4,662	8,729	12,536	16,144	19,680	22,574
	Sample 4	616	2,846	5,447	7,879	10,006	12,034	14,045
	Sample 5	1,149	4,893	9,204	12,972	17,047	20,919	24,274
	Sample 6	1,323	6,449	12,645	18,186	23,821	28,837	33,072
	Sample 7	931	4,704	9,197	12,825	16,863	20,651	23,563
	Sample 8	1,577	7,437	14,210	20,288	26,559	32,259	37,041
	Sample 9	1,072	4,886	9,635	13,953	18,246	21,921	25,377
	Sample 10	1,192	5,469	10,929	16,187	20,751	25,278	29,369
	Sample 11	1,297	6,448	12,820	18,515	24,366	29,199	34,085
	Sample 12	1,420	6,806	13,574	19,397	25,266	30,884	35,725
	Sample 13	651	3,003	5,787	8,189	10,482	12,621	14,498
	Sample 14	1,456	5,990	11,153	16,378	20,988	25,569	29,840
	Sample 15	1,176	4,791	9,015	13,434	16,965	20,554	24,050
	Sample 16	1,168	5,337	10,104	14,639	19,087	22,991	26,522
	Average	1,159	4,890	9,420	13,694	17,647	21,420	24,826
cstacks	40% of individuals	1,718	12,540	22,337	30,287	37,523	44,031	49,695
	60% of individuals	575	6,339	10,795	13,745	16,379	18,582	20,518
	80% of individuals	130	2,426	4,076	4,830	5,520	6,116	6,618

Supporting Information Table S7. Assembly metrics across 80% (r80) of the population for iterations of M0-M11 (the maximum number of mismatches) for the *ETW* dataset for the number of loci assembled, the number of polymorphic loci and the number of SNPs.

	Number of loci	Number of polymorphic loci	Number of SNPs
M0	3,615	120	130
M1	3,094	1,598	2,426
M2	3,227	1,964	4,076
M3	3,216	2,033	4,830
M4	3,220	2,082	5,520
M5	3,231	2,122	6,116
M6	3,244	2,156	6,618
M7	3,261	2,190	7,156
M8	3,266	2,217	7,610
M9	3,253	2,224	8,005
M10	3,254	2,230	8,432
M11	3,261	2,255	8,860

Supporting Information Table S8. Metrics collected for increasing values for M (the maximum number of mismatches) for each dataset (*TRT, PGN* and ETW) for the total number of polymorphic loci across 80% of the population (r80); the number of new r80 loci with each increasing increment of M and the log modulus transformation: (sign(x)*(log10(abs(x)+1)) for plotting these data.

TRT brown trout dataset

	Total r80 loci	No. of new r80 loci	Log modulus transformation
M0/M1	20416	19934	4.30
M1/M2	21528	1112	3.05
M2/M3	21579	51	1.72
M3/M4	21645	66	1.83
M4/M5	21682	37	1.58
M5/M6	21673	-9	-1.00
M6/M7	21519	-154	-2.19
M7/M8	21497	-22	-1.36

PGN king penguin dataset

	Total r80 loci	No. of new r80 loci	Log modulus transformation
M0/M1	6396	6330	3.80
M1/M2	6562	166	2.22
M2/M3	6533	-29	-1.48
M3/M4	6508	-25	-1.41
M4/M5	6508	0	0.00
M5/M6	6513	5	0.78
M6/M7	6526	13	1.15
M7/M8	6523	-3	-0.60

ETW red earthworm dataset

	Total r80 loci	No. of new r80 loci	Log modulus transformation
M0/M1	1598	1478	3.17
M1/M2	1964	366	2.56
M2/M3	2033	69	1.85
M3/M4	2082	49	1.70
M4/M5	2122	40	1.61
M5/M6	2156	34	1.54
M6/M7	2190	34	1.54
M7/M8	2217	27	1.45

Supporting Information Table S9. Metrics for iterating over n (n0-n10) for each of the three datasets: *TRT, PGN* and *ETW* showing results from the script count_fixed_snps.py and assembly metrics from populations at 80% for M2, M4 and M6. Values italicised and in bold represent the value of n that gave the highest amount of polymorphic loci across 80% of the population.

M2	***TRT* brown trout dataset**	n0	n1	***n2***	n3	n4	n5	n6	n7	n8	n9	n10
count_fixed_snps.py	variable sites across the population	67,487	72,663	***74,454***	74,621	74,164	74,244	73,809	73,527	73,453	73,325	73,283
	heterozygous snps in the catalog	68,169	84,783	***98,159***	116,953	137,118	158,233	181,001	204,976	230,649	256,776	283,415
	SNPs across all samples and in the catalog	67,670	77,827	***86,594***	96,014	104,854	113,832	122,649	131,892	142,107	152,179	162,629
	fixed SNPs only in the catalog	183	5,164	***12,140***	21,393	30,690	39,588	48,840	58,365	68,654	78,854	89,346
populations at 80%	Number of loci	60,328	64,270	***64,496***	62,750	61,284	60,081	58,846	57,928	56,952	56,043	55,295
	Number of polymorphic loci	21,483	27,433	***28,040***	27,767	27,518	27,303	27,081	26,977	26,804	26,625	26,581
	Number of SNPs	29,954	39,086	***40,563***	40,321	40,281	40,488	40,444	40,871	41,108	41,446	42,246

M2	***PGN* king penguin dataset**	n0	n1	***n2***	n3	n4	n5	n6	n7	n8	n9	n10
count_fixed_snps.py	variable sites across the population	43,915	54,606	***56,406***	56,568	56,492	56,461	56,494	56,351	56,447	56,457	56,428
	heterozygous snps in the catalog	44,069	76,294	***85,603***	90,837	95,275	99,042	103,101	106,900	110,993	115,125	119,583
	SNPs across all samples and in the catalog	43,963	62,878	***68,783***	71,617	73,476	75,115	76,834	78,108	79,717	81,173	82,433
	fixed SNPs only in the catalog	48	8,272	***12,377***	15,049	16,984	18,654	20,340	21,757	23,270	24,716	26,005
populations at 80%	Number of loci	16,190	21,219	***21,590***	21,510	21,443	21,401	21,365	21,334	21,308	21,268	21,239
	Number of polymorphic loci	6,560	12,475	***12,880***	12,841	12,799	12,776	12,763	12,756	12,756	12,731	12,724
	Number of SNPs	9,170	19,700	***20,951***	20,947	20,904	20,906	20,931	20,985	21,080	21,113	21,148

M2	*ETW* red earthworm dataset	n0	n1	***n2***	n3	n4	n5	n6	n7	n8	n9	n10
count_fixed_snps.py	variable sites across the population	79,147	83,332	***88,967***	89,749	90,116	89,954	89,966	89,949	89,894	89,741	89,928
	heterozygous snps in the catalog	79,296	134,932	***214,585***	302,341	391,988	475,070	553,477	622,428	688,391	750,426	810,244
	SNPs across all samples and in the catalog	79,172	94,928	***122,951***	147,273	168,713	185,498	200,412	212,417	224,486	235,717	246,356
	fixed SNPs only in the catalog	25	11,596	***33,984***	57,524	78,597	95,544	110,446	122,468	134,592	145,976	156,428
populations at 80%	Number of loci	3,230	4,607	***5,109***	4,997	4,948	4,896	4,843	4,792	4,737	4,698	4,673
	Number of polymorphic loci	1,966	3,424	***3,964***	3,926	3,907	3,889	3,863	3,829	3,795	3,775	3,768
	Number of SNPs	4,082	7,534	***9,661***	10,207	10,693	11,105	11,487	11,676	11,855	12,116	12,418

M4	*TRT* brown trout dataset	n0	n1	n2	***n3***	n4	n5	n6	n7	n8	n9	n10
count_fixed_snps.py	variable sites across the population	105,534	108,743	109,896	***110,242***	111,058	110,401	110,219	109,661	109,181	109,250	108,840
	heterozygous snps in the catalog	106,779	121,481	133,258	***143,746***	156,561	175,634	197,648	220,919	246,220	272,221	298,458
	SNPs across all samples and in the catalog	105,779	114,164	121,830	***128,048***	135,552	144,210	154,272	164,363	175,038	186,079	197,448
	fixed SNPs only in the catalog	245	5,421	11,934	***17,806***	24,494	33,809	44,053	54,702	65,857	76,829	88,608
populations at 80%	Number of loci	57,384	61,037	61,151	***60,916***	60,587	59,409	58,236	57,255	56,308	55,471	54,656
	Number of polymorphic loci	21,685	27,390	27,969	***28,061***	27,963	27,780	27,550	27,340	27,169	27,075	26,929
	Number of SNPs	39,489	48,922	49,850	***50,473***	50,365	50,413	49,849	49,914	50,153	50,698	51,079

M4	***PGN* king penguin dataset**	n0	n1	n2	n3	***n4***	n5	n6	n7	n8	n9	n10
count_fixed_snps.py	variable sites across the population	52,646	62,814	64,431	64,542	***65,004***	64,991	65,167	65,101	64,926	64,785	64,963
	heterozygous snps in the catalog	53,120	84,168	92,336	95,762	***99,368***	102,831	106,900	110,659	114,434	118,299	123,010
	SNPs across all samples and in the catalog	52,794	70,867	75,947	77,856	***79,707***	81,310	83,195	84,469	85,869	87,195	89,027
	fixed SNPs only in the catalog	148	8,053	11,516	13,314	***14,703***	16,319	18,028	19,368	20,943	22,410	24,064
populations at 80%	Number of loci	16,045	21,057	21,513	21,559	***21,570***	21,537	21,500	21,463	21,431	21,395	21,365
	Number of polymorphic loci	6,511	12,396	12,882	12,939	***12,952***	12,935	12,913	12,904	12,900	12,873	12,869
	Number of SNPs	9,334	19,998	21,470	21,731	***21,808***	21,799	21,810	21,862	21,947	21,942	22,027

M4	***ETW* red earthworm dataset**	**n0**	n1	n2	n3	n4	***n5***	n6	n7	n8	n9	n10
count_fixed_snps.py	variable sites across the population	152,140	152,480	155,928	160,794	165,475	***166,286***	166,721	166,420	166,368	166,558	166,159
	heterozygous snps in the catalog	152,811	200,547	268,270	339,906	416,194	***494,406***	568,405	636,380	701,322	763,445	762,676
	SNPs across all samples and in the catalog	152,285	164,449	188,227	212,876	237,727	***258,551***	276,520	291,983	306,717	321,361	320,675
	fixed SNPs only in the catalog	145	11,969	32,299	52,082	72,252	***92,265***	109,799	125,563	140,349	154,803	154,516
populations at 80%	Number of loci	3,220	4,777	5,685	6,275	6,599	***6,601***	6,562	6,497	6,446	6,392	6,399
	Number of polymorphic loci	2,083	3,710	4,658	5,260	5,593	***5,621***	5,609	5,560	5,528	5,494	5,501
	Number of SNPs	5,499	10,734	15,135	18,662	21,066	***21,813***	22,293	22,519	22,765	23,048	23,066

M6	*TRT* brown trout dataset	n0	n1	n2	n3	n4	***n5***	n6	n7	n8	n9	n10
count_fixed_snps.py	variable sites across the population	146,353	146,422	147,546	147,671	147,725	***148,034***	147,794	147,373	146,764	146,917	146,498
	heterozygous snps in the catalog	147,897	159,877	171,310	180,966	191,713	***203,413***	215,774	237,989	262,169	287,803	313,842
	SNPs across all samples and in the catalog	146,762	152,091	159,401	165,434	171,585	***178,305***	184,923	195,269	206,132	218,249	230,214
	fixed SNPs only in the catalog	409	5,669	11,855	17,763	23,860	***30,271***	37,129	47,896	59,368	71,332	83,716
populations at 80%	Number of loci	55,524	59,007	58,966	58,746	58,415	***58,215***	57,993	56,985	56,069	55,199	54,357
	Number of polymorphic loci	21,700	27,314	27,877	27,981	27,892	***27,903***	27,843	27,639	27,403	27,345	27,201
	Number of SNPs	49,912	59,325	60,552	61,317	60,847	***61,188***	61,101	60,958	60,321	61,654	61,756

M6	*PGN* king penguin dataset	n0	n1	n2	n3	n4	***n5***	n6	n7	n8	n9	n10
count_fixed_snps.py	variable sites across the population	58,144	67,467	69,072	69,589	69,535	***69,633***	70,117	69,744	69,720	69,784	69,588
	heterozygous snps in the catalog	58,784	88,883	96,962	100,465	103,224	***106,136***	109,317	112,400	116,321	120,323	124,702
	SNPs across all samples and in the catalog	58,307	75,574	80,562	82,897	83,988	***85,506***	87,216	88,058	89,812	91,301	92,840
	fixed SNPs only in the catalog	163	8,107	11,490	13,308	14,453	***15,873***	17,099	18,314	20,092	21,517	23,252
populations at 80%	Number of loci	16,002	20,993	21,440	21,498	21,493	***21,499***	21,495	21,465	21,434	21,406	21,364
	Number of polymorphic loci	6,508	12,367	12,850	12,906	12,913	***12,923***	12,917	12,914	12,907	12,897	12,866
	Number of SNPs	9,505	20,113	21,648	21,841	21,963	***22,079***	22,084	22,114	22,162	22,267	22,180

M6	***ETW* red earthworm dataset**	n0	n1	n2	n3	n4	n5	***n6***	n7	n8	n9	n10
count_fixed_snps.py	variable sites across the population	218,820	216,227	218,315	220,765	225,387	228,966	***232,937***	232,528	232,696	232,928	233,030
	heterozygous snps in the catalog	219,796	263,196	326,373	390,449	459,307	524,986	***589,725***	652,799	715,349	776,091	833,656
	SNPs across all samples and in the catalog	219,141	228,695	250,738	272,224	294,822	315,128	***335,383***	350,604	367,257	383,017	396,838
	fixed SNPs only in the catalog	321	12,468	32,423	51,459	69,435	86,162	***102,446***	118,076	134,561	150,089	163,808
populations at 80%	Number of loci	3,248	4,815	5,834	6,614	7,149	7,478	***7,646***	7,599	7,558	7,498	7,445
	Number of polymorphic loci	2,156	3,801	4,850	5,642	6,188	6,528	***6,698***	6,671	6,643	6,599	6,569
	Number of SNPs	6,630	12,594	18,262	23,507	27,469	30,283	***31,952***	32,326	32,621	32,784	33,008

Supporting Information Table S10. Metrics resulting from testing n=M, n=M-1 and n=M+1 for the optimal value of M for each of the datasets: *TRT* (M5), *PGN* (M2) and *ETW* (M3). Values in italicised and bold represent the value of n that gave the highest amount of polymorphic loci across 80% (*r80*) of the population.

***TRT* brown trout dataset = M5**		**n4**	**n5**	**n6**
populations at 80%	Number of loci	***59,503***	59,240	58,160
	Number of polymorphic loci	***27,980***	27,949	27,723
	Number of SNPs	***55,549***	55,603	55,274
count_fixed_snps.py	variable sites across the population	***129,040***	128,979	128,211
	heterozygous snps in the catalog	***173,332***	185,269	205,933
	SNPs across all samples and in the catalog	***153,104***	159,694	168,784
	fixed SNPs only in the catalog	***24,064***	30,715	40,573
***PGN* king penguin dataset = M2**		**n1**	***n2***	**n3**
populations at 80%	Number of loci	21,219	***21,590***	21,510
	Number of polymorphic loci	12,475	***12,880***	12,841
	Number of SNPs	19,700	***20,951***	20,947
count_fixed_snps.py	variable sites across the population	54,606	***56,406***	56,568
	heterozygous snps in the catalog	76,294	***85,603***	90,837
	SNPs across all samples and in the catalog	62,878	***68,783***	71,617
	fixed SNPs only in the catalog	8,272	***12,377***	15,049
***ETW* red earthworm dataset = M3**		**n2**	***n3***	**n4**
populations at 80%	Number of loci	5,501	***5,911***	5,881
	Number of polymorphic loci	4,432	***4,860***	4,860
	Number of SNPs	12,894	***15,302***	15,922
count_fixed_snps.py	variable sites across the population	123,208	***128,483***	129,284
	heterozygous snps in the catalog	240,158	***317,281***	402,824
	SNPs across all samples and in the catalog	155,973	***182,814***	205,994
	fixed SNPs only in the catalog	32,765	***54,331***	76,710

Supporting Information Table S11. Raw results for the 16 samples for the *Stacks* default parameters (m3, M2, n1) and the optimal parameters for the three datasets: *TRT* (m3, M5, n4); *PGN* (m3, M2, n2); *ETW* (m3, M3, n3) for (a) the number of assembled loci, (b) the number of polymorphic loci and (c) the number of SNPs.

(a) Number of assembled loci

		Default *TRT*	**Optimal *TRT***	**Default *PGN***	**Optimal *PGN***	**Default *ETW***	**Optimal *ETW***
ustacks	Sample 1	91,141	89,066	66,890	66,890	111,420	110,533
	Sample 2	91,994	89,897	55,667	55,667	121,132	120,170
	Sample 3	92,814	90,723	24,942	24,942	126,816	125,856
	Sample 4	120,793	118,613	41,468	41,468	80,171	79,564
	Sample 5	99,023	97,207	30,297	30,297	136,641	135,669
	Sample 6	96,681	94,498	57,860	57,860	179,720	178,241
	Sample 7	93,397	91,266	58,080	58,080	133,928	132,887
	Sample 8	86,839	84,932	59,034	59,034	184,066	182,150
	Sample 9	89,548	87,552	58,718	58,718	141,169	140,014
	Sample 10	95,873	93,717	44,617	44,617	158,787	157,407
	Sample 11	91,059	88,979	66,064	66,064	184,909	183,372
	Sample 12	87,805	85,820	65,525	65,525	193,807	192,198
	Sample 13	78,352	76,638	75,713	75,713	82,593	81,966
	Sample 14	98,438	96,269	92,793	92,793	157,285	155,955
	Sample 15	92,438	90,284	102,434	102,434	132,098	130,979
	Sample 16	94,134	92,001	91,300	91,300	146,248	145,051
	Average	92,626	90,504	58,876	58,876	138,905	137,842
	Total	278,177	244,027	162,004	154,982	960,690	865,474
cstacks	40% of individuals	85,341	80,224	70,498	71,067	47,712	55,061
	60% of individuals	76,604	71,886	48,008	48,728	15,031	18,997
	80% of individuals	64,272	59,482	21,220	21,584	4,606	5,912

(b) Number of polymorphic loci

		Default *TRT*	Optimal *TRT*	Default *PGN*	Optimal *PGN*	Default *ETW*	Optimal *ETW*
ustacks	Sample 1	8,311	10,162	4,141	4,147	7,709	5,649
	Sample 2	8,881	10,664	2,809	2,810	8,120	6,032
	Sample 3	8,633	10,391	671	670	8,781	6,421
	Sample 4	14,834	16,968	1,892	1,889	5,477	4,055
	Sample 5	11,335	13,208	924	925	9,293	6,645
	Sample 6	9,003	10,848	2,993	2,994	12,499	9,140
	Sample 7	8,933	10,840	3,090	3,092	9,242	6,690
	Sample 8	7,550	9,176	3,245	3,253	14,091	10,492
	Sample 9	7,712	9,443	4,175	4,179	9,622	7,041
	Sample 10	9,066	10,946	4,241	4,236	11,011	8,012
	Sample 11	7,766	9,625	5,673	5,694	12,717	9,283
	Sample 12	7,656	9,496	7,526	7,531	13,539	9,838
	Sample 13	5,909	7,708	6,888	6,875	5,750	4,299
	Sample 14	9,042	11,007	11,395	11,407	11,164	8,347
	Sample 15	8,140	10,184	13,167	13,144	9,026	6,777
	Sample 16	8,499	10,400	8,715	8,702	10,138	7,399
	Average	8,566	10,396	4,158	4,163	9,458	6,909
	Total	48,963	54,283	47,977	49,150	78,413	112,011
cstacks	40% of individuals	37,323	39,577	36,568	37,677	21,840	32,650
	60% of individuals	33,496	34,981	26,542	27,442	9,611	14,238
	80% of individuals	27,398	27,928	12,475	12,872	3,423	4,860

(c) Number of SNPs

		Default *TRT*	Optimal *TRT*	Default *PGN*	Optimal *PGN*	Default *ETW*	Optimal *ETW*
ustacks	Sample 1	13,682	26,656	5,544	5,573	7,709	11,026
	Sample 2	14,409	27,217	3,913	3,937	8,120	11,919
	Sample 3	14,448	27,425	877	880	8,781	12,623
	Sample 4	22,256	39,282	2,557	2,545	5,477	7,774
	Sample 5	17,219	31,228	1,290	1,313	9,293	13,012
	Sample 6	14,747	28,001	4,204	4,161	12,499	18,126
	Sample 7	14,561	28,262	4,303	4,318	9,242	12,878
	Sample 8	12,378	24,142	4,454	4,469	14,091	20,443
	Sample 9	12,704	24,983	5,418	5,400	9,622	13,935
	Sample 10	14,853	28,168	5,591	5,518	11,011	16,153
	Sample 11	12,850	25,276	7,300	7,350	12,717	18,566
	Sample 12	12,807	25,344	9,818	9,820	13,539	19,517
	Sample 13	9,995	20,862	9,067	9,089	5,750	8,204
	Sample 14	14,812	28,967	14,752	14,728	11,164	16,460
	Sample 15	13,575	26,880	17,364	17,311	9,026	13,347
	Sample 16	14,059	27,351	11,829	11,713	10,138	14,812
	Average	14,234	27,284	5,481	5,459	9,458	13,641
	Total	84,783	172,765	76,294	85,741	134,932	316,983
cstacks	40% of individuals	59,736	94,788	56,514	61,091	44,343	99,781
	60% of individuals	50,722	76,711	41,418	44,613	21,024	45,742
	80% of individuals	38,894	55,170	19,686	20,903	7,532	15,308

Supporting Information Table S12. Standard population genetics statistics: FIS; HE and FST calculated for each of the three datasets at 80% across the population (*TRT*, *PGN* and *ETW*) comparing default parameters (m3, M2, n1) to the optimal set: *TRT* = m3, M5, n4; *PGN* = m3, M2, n2; *ETW* = m3, M3, n3. FST was calculated between the two populations for the *TRT* and *PGN* datasets, and for these two datasets, FIS and HE were calculated per population.

	***TRT* default**		***TRT* optimal**		***PGN* default**		***PGN* optimal**		***ETW* default**	***ETW* optimal**
	Clean	***Metal***	***Clean***	***Metal***	***PCM***	***KER***	***PCM***	***KER***	***OL2***	
FIS	0.118	0.082	0.092	0.043	0.367	0.160	0.406	0.171	0.283	0.430
HE	0.206	0.170	0.204	0.180	0.182	0.216	0.191	0.220	0.220	0.245
FST	0.060		0.054		0.049		0.050		N.A	N.A

Supporting Information Table S13. Global SNP diversity statistics: nucleotide (Pi); gene and haplotype diversity calculated for each of the three datasets (*TRT*, *PGN* and *ETW*) with increasing values for the mismatch parameter (M) from M1-M6.

	TRT			***PGN***			***ETW***		
	Nucleotide (Pi) diversity	**Gene diversity**	**Haplotype diversity**	**Nucleotide (Pi) diversity**	**Gene diversity**	**Haplotype diversity**	**Nucleotide (Pi) diversity**	**Gene diversity**	**Haplotype diversity**
M1	0.314	0.329	0.351	0.247	0.269	0.279	0.472	0.489	0.610
M2	0.337	0.351	0.487	0.283	0.290	0.371	0.468	0.498	0.799
M3	0.355	0.362	0.595	0.303	0.297	0.427	0.474	0.504	0.990
M4	0.369	0.372	0.711	0.318	0.300	0.472	0.480	0.511	1.159
M5	0.383	0.378	0.834	0.328	0.303	0.502	0.487	0.516	1.325
M6	0.390	0.384	0.950	0.338	0.304	0.526	0.491	0.519	1.462

Supporting Information Table S14. The number of consensus loci that aligned against the respective reference genomes using *ref map* and the *de novo - integrated* method for: *TRT* (*Salmo salar* aligned to *Salmo salar*); *PGN* (*Aptenodytes patagonicus* aligned to *Aptenodytes forsteri*) and *ETW* (*Lumbricus rubellus* aligned to *Lumbricus rubellus*). Unique positions represent cases where one consensus locus aligned to one position in the genome; multiple positions represent situations where multiple consensus loci aligned to the same position in the genome.

***TRT* brown trout dataset**		
Ref map		66,806
De novo - Integrated		116,537
	Unique positions	84,532
	Multiple positions	32,005
***PGN* king penguin dataset**		
Ref map		91,902
De novo - Integrated		117,783
	Unique positions	104,793
	Multiple positions	12,990
***ETW* red earthworm dataset**		
Ref map		100,989
De novo - Integrated		253,681
	Unique positions	172,209
	Multiple positions	81,472

Supporting Information Table S15. Raw data for the frequency of the number of SNPs found per catalog locus for different iterations of the n parameter (the number of mismatches allowed between stacks in the *catalog*) for (a) *catalog* loci SNP frequency (b) sample loci SNP frequency across 80% of the population (*r80 loci*) for the datasets *TRT*, *PGN* and *ETW*.

***TRT* brown trout dataset (M2) – *catalog* SNP frequency**

	n0	n1	n2	n3	n4	n5	n6	n7	n8	n9	n10	*Ref map*
1	25684	31190	28910	27064	25688	24573	23677	22888	22166	21483	20903	17064
2	9331	10929	12252	11019	10260	9674	9255	8905	8612	8329	8094	5911
3	2456	3311	4146	6472	5867	5396	5033	4815	4617	4430	4295	2540
4	1112	1286	1715	2670	4692	4229	3900	3676	3495	3357	3240	1492
5	681	708	842	1358	2165	3959	3608	3319	3135	3003	2932	952
6	424	469	576	826	1308	1973	3567	3288	3057	2882	2738	694
7	289	326	385	514	735	1111	1760	3246	2941	2709	2563	502
8	145	212	267	328	483	725	1018	1596	3041	2812	2612	378
9	114	156	226	270	379	524	728	995	1540	2904	2705	285
10	73	118	158	186	297	400	570	748	986	1389	2636	214

***TRT* brown trout dataset (M2) – sample SNP frequency (*r80 loci*)**

	n0	n1	n2	n3	n4	n5	n6	n7	n8	n9	n10	*Ref map*
1	15112	18847	18709	18493	18274	18070	17903	17776	17626	17471	17365	12455
2	5078	6643	7144	6995	6911	6798	6716	6658	6595	6530	6503	3735
3	901	1400	1603	1667	1631	1613	1586	1574	1545	1520	1501	1365
4	199	286	365	395	451	447	442	436	435	415	432	721
5	78	98	108	123	134	198	189	190	197	205	208	428
6	56	72	56	47	54	97	142	158	148	148	148	306
7	34	46	32	29	36	39	62	103	106	108	105	254
8	9	18	14	10	12	27	28	59	100	100	95	175
9	13	14	6	4	6	8	6	14	34	80	94	133
10	1	7	2	3	7	4	6	6	12	33	89	91

***PGN* king penguin dataset (M2) – *catalog* SNP frequency**

	n0	n1	n2	n3	n4	n5	n6	n7	n8	n9	n10	*Ref map*
1	20163	30802	28781	28120	27795	27595	27444	27308	27188	27083	26965	24675
2	7019	11475	12735	12171	11959	11812	11720	11626	11593	11527	11479	10099
3	1481	3636	4602	5203	5036	4928	4852	4821	4772	4748	4711	3865
4	453	1103	1575	1844	2255	2151	2074	2030	2014	1975	1962	1394
5	224	391	589	761	936	1218	1168	1123	1083	1078	1043	604
6	109	174	257	366	478	590	875	820	782	756	731	306
7	96	129	171	220	291	372	437	708	660	623	609	191
8	52	80	115	135	145	199	260	357	580	557	530	154
9	33	41	60	94	106	131	184	252	319	540	512	125
10	18	33	39	72	84	111	146	193	223	274	488	100

***PGN* king penguin dataset (M2) – sample SNP frequency (*r80 loci*)**

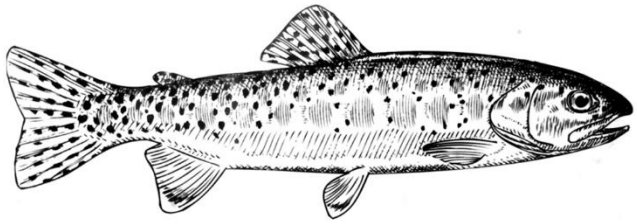
	n0	n1	n2	n3	n4	n5	n6	n7	n8	n9	n10	*Ref map*
1	4571	7357	7303	7267	7246	7232	7222	7214	7202	7188	7183	7649
2	1508	3488	3704	3679	3660	3650	3641	3635	3634	3623	3618	3799
3	378	1263	1410	1418	1411	1404	1403	1396	1398	1391	1388	1614
4	78	292	354	362	356	353	353	355	351	351	351	624
5	19	57	81	84	91	97	94	97	97	94	95	256
6	2	8	18	22	22	27	26	27	30	28	29	121
7	2	6	6	6	11	8	17	18	17	17	17	73
8	2	1	2	2	1	1	3	7	19	19	19	43
9	0	3	1	0	1	2	3	6	5	12	10	46
10	0	0	1	1	0	1	1	1	2	5	8	49

***ETW* red earthworm dataset (M2) – *catalog* SNP frequency**

	n0	n1	n2	n3	n4	n5	n6	n7	n8	n9	n10	*Ref map*
1	19128	50904	43683	38627	35668	33776	32452	31461	30658	30002	29444	9341
2	13308	15116	32752	27199	24271	22365	21197	20385	19808	19355	18923	7876
3	4105	5931	9917	23605	20679	18620	17482	16640	16065	15621	15285	7736
4	2580	3298	5637	8432	18878	16727	15437	14542	13926	13482	13133	7716
5	801	1307	2549	5052	6817	15101	13803	12846	12209	11736	11415	6929
6	432	635	1345	2846	4720	5942	12420	11471	10861	10357	10022	6503
7	218	354	737	1574	3088	4398	5276	10385	9771	9302	8895	5791
8	127	254	444	1018	1888	3133	4021	4658	8970	8414	8040	4825
9	80	151	267	606	1212	2008	2967	3630	4122	7695	7354	4015
10	37	122	184	382	771	1356	2010	2663	3149	3574	6708	3299

***ETW* red earthworm dataset (M2) – sample SNP frequency (*r80 loci*)**

	n0	n1	n2	n3	n4	n5	n6	n7	n8	n9	n10	*Ref map*
1	822	1301	1261	1183	1157	1126	1094	1074	1061	1050	1042	899
2	611	979	1109	1015	968	927	911	893	875	858	851	859
3	286	637	785	777	729	706	683	671	652	645	633	800
4	146	309	454	470	472	465	447	434	428	417	415	825
5	58	126	209	258	285	306	292	280	275	269	266	736
6	21	36	86	124	135	144	168	162	155	156	152	717
7	11	18	38	59	77	93	95	109	109	107	102	655
8	4	10	16	28	41	57	70	80	86	84	84	617
9	4	5	4	8	19	29	42	55	63	67	69	465
10	2	2	2	4	18	22	30	34	37	41	49	446



Chapter IV

Rapid and repeated genomic adaptations to unique cocktails of metal contaminants in brown trout

Manuscript in preparation

AUTHORS: J.R. Paris[1], R.A. King[1], K. Moore[2], J.M. Catchen[3] & J.R. Stevens[1]

[1] Biosciences, College of Life and Environmental Sciences, University of Exeter, Exeter, UK

[2] Exeter Sequencing Service, Geoffrey Pope Building, University of Exeter, Exeter, EX4 4QD

[3] Department of Animal Biology, University of Illinois at Urbana–Champaign, Urbana, IL 61801

J.R. Paris designed the research, prepared the RAD libraries, conducted the data analysis and wrote the manuscript

R.A. King designed the research, prepared the RAD libraries, assisted in data analysis and improved the manuscript

K. Moore prepared the RAD libraries

J.M. Catchen assisted with and contributed to data analysis

J.R. Stevens designed the research and improved the manuscript

Rapid and repeated local adaptation of brown trout populations to unique cocktails of metal contaminants

AUTHORS: J.R. Paris[1], R.A. King[1], K.A. Moore[2], J.M Catchen[3] and J.R. Stevens[1]

Manuscript in preparation

[1] Biosciences, College of Life and Environmental Science, Geoffrey Pope Building, University of Exeter, Exeter, EX4 4QD

[2] Exeter Sequencing Service, Geoffrey Pope Building, University of Exeter, Exeter, EX4 4QD

[3] Department of Animal Biology, University of Illinois at Urbana–Champaign, Urbana, IL 61801

ABSTRACT

Human-altered environments are extremely powerful in shaping the demographics of natural populations. Brown trout (*Salmo trutta* L.) occupying rivers impacted by historical mining activity offered a unique opportunity to assess distinct patterns of local adaptation to a patchwork of metal polluted rivers. A previous study of genetic variation in these fish using microsatellite loci, identified marked differences between populations inhabiting metal-impacted rivers, compared to fish from clean reference rivers. However, no patterns of selection were identified. Consequently, further exploration of the effects of continuing metal pollution in driving differences between populations of trout inhabiting metal-impacted and so-called 'clean' rivers are required to fully understand the genetic basis of the ability of these trout to thrive in waters known to be lethal to metal-naïve fish. Accordingly, we analysed five populations of brown trout (*Salmo trutta L.*) from across west Cornwall, UK, in order to identify patterns driving divergence based on metal-tolerance. Via restriction-site-associated DNA sequencing (RADseq), we used a set of 5,825 polymorphic loci to explore the patterns and processes conferring local adaptation in these metal-tolerant trout populations. We observed a recapitulation of the previous structure identified in these fish using neutral microsatellite markers. Latent fixed factor models (LFMM) of covarying markers showed that twenty-nine loci were associated with metal pollutant data, and one

locus was correlated specifically with arsenic, cadmium, nickel and zinc. Patterns of isolation-by-environment (IBE) showed that genetic differentiation between populations was significantly correlated with cadmium, nickel and zinc. Specifically, the identification of loci associated with the ability to cope with metal pollutants allows the relative importance of adaptation to metal pollution in these fish to be evaluated. This study demonstrates the remarkable ability of fish populations to undergo rapid local adaptation to human-altered environments, driven by unique cocktails of metal pollutants in neighbouring fish populations.

KEYWORDS: local adaptation, brown trout, RAD-seq, selection, genome, LFMM, F_{ST}

INTRODUCTION

Selection is a key factor in shaping the historical and contemporary demographics of natural populations. Traditionally, it was believed that processes of natural selection act over many generations and across relatively large geographic areas (Kinnison & Hendry 2001; Hendry & Kinnison 2001; Reznick & Ghalambor 2001; Hairston *et al.* 2005). However, recognition of the potency of local adaptation to operate at both small spatial and temporal scales (Antonovics 1971; Polyakov *et al.* 2004; Whitehead *et al.* 2017), and a more enlightened appreciation of the intricacy of natural selection, have now emerged (Messer *et al.* 2016; Hadfield 2016; Hermisson & Pennings 2017). To distinguish local adaptation from selection more generally, it is important to recognise how local selection pressures can drive differential allele frequencies between neighbouring populations, even when these populations are connected by gene flow (Williams 1996; Kawecki & Ebert 2004; Blanquart *et al.* 2013; Savolanien *et al.* 2013). Divergent selection on ecological traits is powerful in shifting the mean allele frequencies of populations, resulting in higher fitness of a population to the local ecology of their surrounding environment (Orr 1998; Stortz 2005; Morjan & Rieseberg 2004).

Salmonids are a group of teleost fish, well recognised for their propensity to form locally adapted and distinctive populations (Taylor 1991; Keeley *et al.* 2007; Fraser *et al.* 2011), which in some cases can lead to significant divergence of within-species 'ecotypes'. For example, in North America, planktivorous and piscivorous ecotypes of lake trout (*Salvelinus namaycush*) are recognised (Goetz *et al.* 2010; Perreault-Payette *et al.* 2017), ferox trout are a piscivorous form of brown trout (*Salmo trutta*) found in glacial lakes of Northern Britain (Fergurson & Mason 1981; Duguid *et al.* 2006); and lake whitefish (*Coregonus clupeaformis*) are known to undergo ecotype divergence based on limnetic and benthic ecological niches (Campbell & Bernatchez 2004; Landry *et al* 2007). Moreover, the time-scales over which new ecotypes may arise as a result of local adaptation to novel environments can be rapid. In other teleost fish, such as the three-spine stickleback (*Gasterosteus aculeatus*), migratory lake and resident stream ecotypes have diverged in ~150 years (Marques *et al.* 2016), and similarly, rapid colonisation of new freshwater

habitats in this species appears to have occurred in 50 years, after an earthquake resulted in availability of new niche space (Lescak *et al.* 2015).

One of the most powerful forces affecting species', both globally and at a local scale, is the activity of humans (Vitousek *et al.* 1997; Crutzen 2002; Allendorf & Hard 2009; Steffen *et al.* 2011). In myriad ways, humans have placed an unprecedented pressure on populations to quickly evolve and adapt (Hamilton *et al.* 2016; Reid *et al.* 2016; Kennedy *et al.* 2016) or risk becoming locally extinct (Lande 1998; De La Vega-Salazar *et al.* 2003; Colson & Hughes 2004). Understanding the evolutionary forces and the genomic basis of such adaptations remains a primary focus in contemporary evolutionary and conservation biology (Stockwell *et al.* 2003; Allendorf *et al.* 2010; Valladares *et al.* 2014).

Aquatic environments across Britain have been moulded by a legacy of industrial activity and associated pollution; in southwest England, this largely took the form of metal mining, resulting in highly metal-contaminated rivers across the region. Despite chronic metal pollution and the ecological decimation of many of these aquatic environments (e.g. Brown 1977; Bryan & Gibbs 1983; Environment Agency 2013), brown trout (*Salmo trutta* L.) populations are known to reside in these extremely metal contaminated rivers, and, moreover, do not appear to display reduced fitness in these challenging environments. Previous research points to the role of rapid local adaptation in conferring metal-tolerance in these fish (Uren-Webster *et al.* 2013; Minghetti *et al.* 2014; Paris Chapter 5 & Chapter 6). In particular, a previous study using 23 microsatellite loci showed that, despite distinct differences in patterns of neutral genetic diversity between metal-impacted and clean-river fish populations, loci under selection could not be discerned (Paris *et al.* 2015). Importantly, the observed patterns suggested that metal pollution is driving not only differences between metal-impacted and clean-river populations, but also that each metal-impacted group may be locally adapted to the unique 'cocktail' of metals they experience in their native river.

Quantifying local adaptation has proven to be a challenging task. Traditionally, it has been assessed through reciprocal transplant experiments, which remain a powerful method in relating the fitness of local populations to their home environment (Hereford 2009; Westley *et al.* 2012; Bougas *et al.* 2016). The

continued development of genetic markers means that population genetic approaches are now the most commonly used methods for exploring the processes and consequences of local adaptation. Such methods typically involved a "bottom-up" approach in assessing local adaptation and are somewhat limited by requiring an *a priori* identification of genes of interest and the use of species-specific genetic resources (Lee & Mitchell-Olds 2006; Larsen *et al.* 2007; Grivet *et al.* 2011). Recently, developments in the generation of cost-effective reduced-representation genomic data (e.g. restriction-site-associated DNA sequencing – RADseq) have proven to offer a promising approach for detecting the genome-wide signals responsible for adaptive differentiation (Davey *et al.* 2011; Savolanien *et al.* 2013; Andrews *et al.* 2016).

Despite the ability to screen for thousands of genetic markers, only a small proportion of these will be putatively under selection (Hess *et al.* 2013; Narum *et al.* 2013) and unpicking the significance of these signals to empirically support local adaptation remains a challenge (Tiffin & Ross-Ibarra 2014; Hoban *et al.* 2016). More recently, significant interest has focused on the mechanisms of selection, such as whether adaptation to novel environments is based on standing genetic variation, or whether it arises through *de novo* mutation (Barrett & Schluter 2008; Elmer & Meyer 2001), and also whether it is driven by the effects of one or a few loci of large effect (i.e. hard sweeps), versus the role of multiple loci of small effect (i.e. polygenic soft sweeps) (Pritchard & di Renzo 2010; Messer & Petrov 2013; Bernatchez *et al.* 2016). Assessing the relative importance of these mechanisms is paramount in exploring the genomic processes underlying selection contributing to local adaptation.

The overall aim of this study was to investigate local adaptation in contributing to metal-tolerance in brown trout populations living in extremely polluted rivers in southwest England. Using RADseq, we assessed the value of genome-wide polymorphic loci in providing evidence of local adaptation within two metal-tolerant populations of trout, and compared these to three populations of trout from nearby uncontaminated ('clean') rivers. Specifically, we aimed to assess (i) whether local adaptation can be evidenced in metal-tolerant trout; (ii) whether local adaptation in different metal-tolerant trout populations has evolved independently and (iii) the underlying selection processes contributing to local adaptation Our investigation illustrates the capacity of rapid and repeated local

adaptation in conferring unique patterns of metal-tolerance in brown trout populations, ultimately showcasing the species' resilience to detrimental anthropogenic change.

MATERIALS AND METHODS

Populations and sampling

We sampled 100 resident brown trout from five rivers across the southwest of England (Figure 1). Due to the underlying metalliferous geology, and the likelihood that mining historically occurred over much of the landscape, it is recognised that virtually all the rivers in this region have been metal-impacted in some way. We designated comparatively 'clean' reference rivers on the basis of contemporary metal contamination levels (data provided by the Environment Agency; Table 1). On the basis of neutral genetic markers, populations of trout from these clean rivers have been shown to comprise similar genetic profiles (Paris *et al.* 2015; King *et al.* 2016). We sampled trout from three clean rivers: Camel, Carrick Roads and Gannel and two metal-impacted rivers: Hayle and Red River. Importantly, the metal-impacted rivers are both affected by high concentrations of metals, but in comparison to one other, comprise different "cocktails" of metal pollution. For example, the River Hayle experiences metal pollution primarily comprised of elevated concentrations of nickel (Ni) and zinc (Zn), whereas the Red River contains high concentrations of arsenic (As) and iron (Fe) – Table 1.

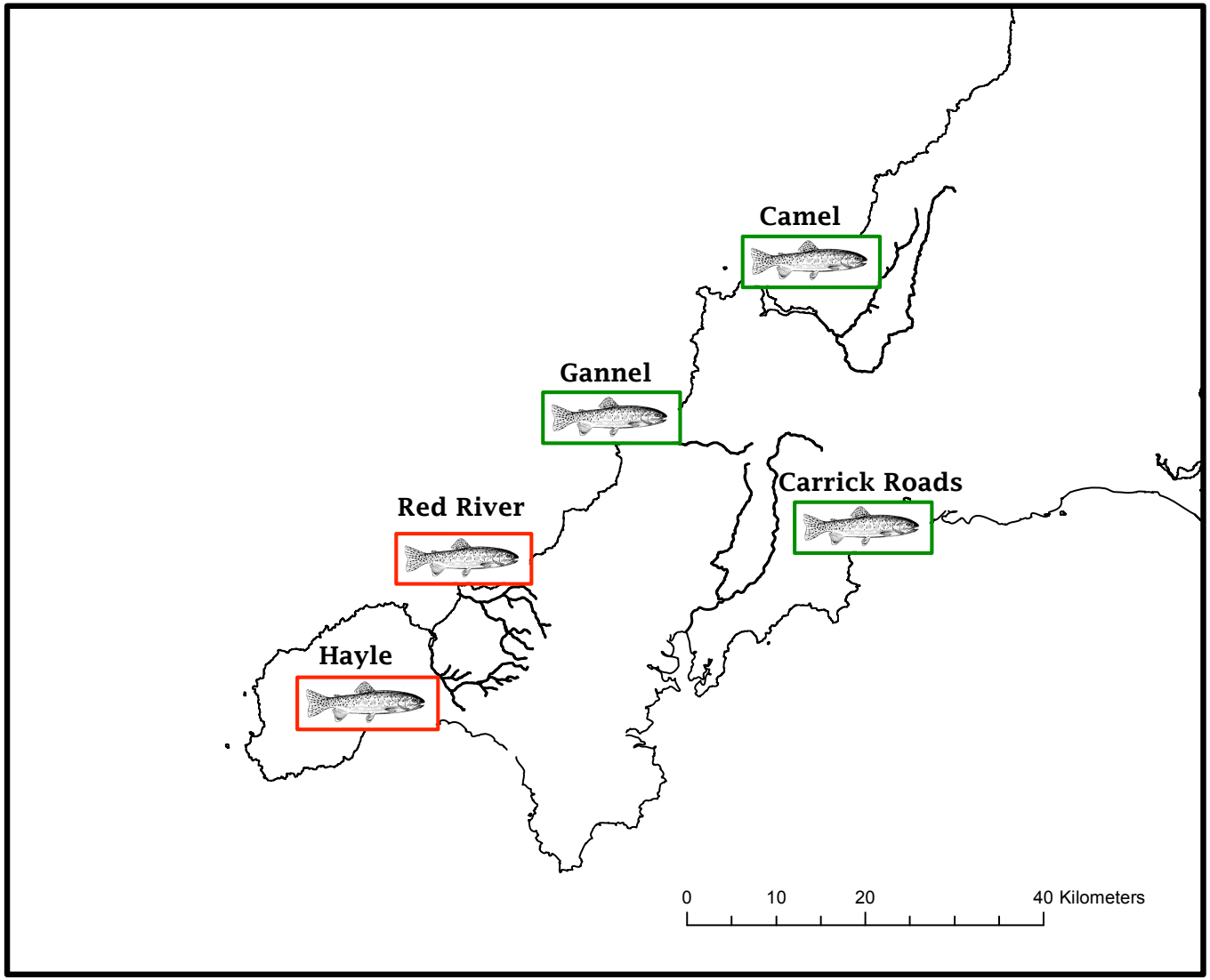


Figure 1. Map of sampled rivers in the southwest of England. Red boxes indicate populations sampled from metal-impacted rivers and green boxes indicate populations sampled from comparatively clean reference rivers.

Table 1. Populations, sampled individuals and metal concentrations for the five rivers chosen for analysis. Three rivers were sampled as representing 'clean' groups (total n=60) and two rivers were sampled for trout representing differentially-impacted metal groups (total n=40). Concentrations of metals are represented as µg/L for arsenic (As), cadmium (Cd), copper (Cu), iron (Fe), nickel (Ni) and zinc (Zn) and are long-term averages over a 15-year period (2000-2014).

Population	*Group*	*N*	*As*	*Cd*	*Cu*	*Fe*	*Ni*	*Zn*
Camel	Clean	20	4.85	0.2	4.06	309.91	11.41	72.82
Carrick Roads	Clean	20	5.85	0.48	3.7	755.38	8.55	103.59
Gannel	Clean	20	4.55	0.3	6	1268.58	3.92	27.84
Hayle	Metal	20	12.84	1.24	63.5	388.61	25.53	555.48
Red River	Metal	20	73.2	1.08	87.1	3395.89	15.17	394.78

Trout from each of the rivers were sampled between 2011 and 2014. Fish were caught by electrofishing under Environment Agency authorisation, with assistance from the Westcountry Rivers Trust. Fish were anesthetised using either MS-222, clove oil or benzocaine (10 g/100 mL ethanol) diluted 1:2000 in river water prior to fin-clip removal. Fin-clips were stored in 95% ethanol in a cold room at 4°C prior to DNA extraction. We chose the largest trout (based on fork-length > 150mm) for inclusion in the RADseq analysis as fish of this size were most likely to represent river-resident adults or sub-adults, thereby removing potential bias by sampling non-river residents and anadromous sea trout.

DNA extraction and RADseq library preparation

DNA was extracted from adipose fin-clips using the QIAGEN DNeasy Blood & Tissue Kit (QIAGEN, Hilden, Germany). We used a combination of the RAD-seq library preparation protocols of Etter *et al.* (2011) and Miller *et al.* (2012). DNA was quantified using the QuantiFluor ONE dsDNA assay system (Promega, Wisconson, USA) and subsequently normalised to 500 ng in 40 µl. Briefly, approximately 500 ng of genomic DNA of each sample was digested with *SbfI*. Individually barcoded P1 adapters were ligated onto the *SbfI* cut site. Individuals were multiplexed and DNA was sheared using either a Bioruptor UCD-200 or Covaris E220. Fragments were size selected for approximately 300-600bp using AmpureXP beads. Fragments were blunt-end repaired and 3'-dA overhangs were added before addition of the P2 adapter to the 3' end. PCR was carried out in a BIO-RAD MyCycler Thermal Cycler in 3 x 50µl reaction volumes containing 2 µL of library, 2 µL each of P1 and P2 adapter primer (10 uM), 25 µL of 2X Phusion Master Mix (New England Biolabs, Massachusetts, USA) and 19 µL of RNase-Free Water. We followed the PCR conditions of Miller *et al*. (2012), using 12 cycles of amplification to reduce PCR duplicates. Individual libraries were randomly pooled and multiplexed into libraries of between 24 to 36 samples per Illumina lane. We sequenced libraries as both single-end (125bp Rapid Run Mode) and paired-end (100bp Standard Mode) on the Illumina HiSeq 2500 at the University of Exeter Sequencing Service.

Data processing and RAD loci construction

RADseq data was processed and analysed using Stacks v 1.42 (Catchen *et al.* 2011; 2013). Raw reads were cleaned and demultiplexed using `process_radtags`, discarding reads with quality scores below 10 and trimming all reads to 90bp to remove and low-quality 3' bases. Optimised parameter sets for construction of RAD loci *de novo* (using `denovo_map`) were selected using the method of Paris *et al.* (2017), which showed the optimal parameters to be a minimum read depth of five (m5), a maximum number of 5 mismatches between alleles (M5) and a distance of four nucleotides for catalog construction (n4). Due to the ancient whole genome duplication (WGD) event in salmonids, we were cautious of merging paralogous loci, and so the maximum number of loci per individual (--max_locus_stacks) was specified as 2. We first built a *de novo* catalog using all individuals in the dataset. Coverage of the samples in this catalog was assessed, and the top 75 samples (with coverage ≥15x) were used to build a second catalog, to which all samples were mapped back using `sstacks`. The `populations` module was run on all samples from the five populations, specifying that a locus had to be present in every population (-p 5), across 80% of the samples within each population (-r 0.8), with a minor allele frequency (MAF) of 10% (--min_maf 0.1) and a maximum observed heterozygosity for each locus no more than 70% (--max_obs_het 0.7). We exported genotypic data in Genepop format (Raymond & Rousset 1995) and used PGD Spider v. 2 (Lischer & Excoffier 2012) to convert the genepop file for subsequent analyses.

Population genetics statistics

Population genetic statistics of diversity and differentiation were calculated in Genodive v. 2.0b27 (Meirmans & van Tienderen 2004). Expected heterozygosity (H_E), observed heterozygosity (H_O) and the inbreeding coefficient (F_{IS}) were calculated for each of the five populations individually based on 10000 permutations of the dataset. We also computed the global F_{ST} across the whole dataset, and pairwise F_{ST} between each pair of populations, using the method of Weir & Cockerham (1984). To assess the relationship between genetic differentiation and geographic distance, isolation-by-distance (IBD) was assessed using a Mantel test. Geographic distance was calculated

using the coastal distance (in km) between each river mouth. We also explored patterns of isolation-by-environment (IBE) by using Mantel tests investigating relationships between genetic differentiation and the metal pollution of each river. Euclidian distances were calculated based on the concentration (in µg/L) of each metal within each river (as shown in Table 1) using the ecodist package in R (Goslee & Urban 2007).

Two AMOVAs were conducted to assess the relative contribution of river basin and metal contamination in driving genetic differentiation. The first AMOVA specified each of the five populations nested within either a clean or a metal group (Clean and Metal), and the second specified each of the five populations nested within a clean group, and either one of the two differentially metal-impacted groups (Clean, Hayle, Red River). AMOVAs were implemented based on 10000 permutations of the dataset, using the infinite alleles model (IAM) in Genodive. Finally, the effective population size (N_e) of each population was calculated using the co-ancestry method of Nomura *et al.* (2008) as implemented in NeEstimator v. 2.01 (Do *et al.* 2014) using the linkage disequilibrium model under a random mating scenario, using 0.01 as the lowest allele frequency for the critical value cut-off.

Population structure

To assess potential population structure and genetic distinctiveness between the five populations, three methods were used: Structure (Prichard *et al.* 2000), a discriminant analysis of principal components (DAPC; Jombart *et al.* 2010) and a neighbour-joining dendrogram. Structure attempts to cluster individuals based on minimising Hardy-Weinberg equilibrium and gametic disequilibrium, but may sometimes fail in discerning more complex population structure (Jombart *et al.* 2010). DAPC is a multivariate analysis that may be more efficient in identifying such hierarchical structure. Further, a neighbour-joining dendrogram was constructed in order to explore the genetic relationships between each of the populations. All population structure analyses were assessed using a single random SNP from each polymorphic locus.

The command line version of Structure was used (Prichard *et al.* 2000), automating the process using the program StrAuto 1.0 (Chhatre & Emerson 2017), running the full dataset for 10 runs of each K of K=1 through to K=10,

using a burn-in of 10000, before collecting data based on 10000 MCMC iterations. The optimal value of K was assessed using the ΔK statistic (Evanno *et al.* 2005). Structure plots were visualised using the pophelper package implemented in R (Francis 2017).

To unpick possible fine-scale structure between populations, we used DAPC in the R package adegenet (Jombart 2008; Jombart *et al.* 2011). We retained the number of PCs that allowed for a description of the most variance between groups (>90% of conserved variance) and retained the first three discriminant functions for analysis. Eigenvalues were used to assess the most important principal components in describing the multi-dimensional space of the genetic relationships.

A neighbour-joining dendrogram was constructed using the aboot function in the R package poppr (Kamvar *et al.* 2014; 2015). We used Nei's genetic distance (Nei 1972) to calculate the genetic 'distance' between samples. The dendrogram was constructed using a random seed of 999, a cut-off value of 50 to return bootstrap values on each node, replicating 1000 bootstraps over the dataset.

De novo loci under divergent and balancing selection

To distinguish between loci potentially under selection amongst the five populations, two different approaches were used: outlier methods that detect loci as extreme F_{ST} outliers, and Latent Factor Mixed Models (LFMM; Frichot *et al.* 2013) to detect covarying loci of small effect associated with metal pollution.

Due to differences in implementation and the potentially high false positive rate of different outlier approaches (Excoffier *et al.* 2009; Narum & Hess 2011; De Mita *et al.* 2013), two methods to detect F_{ST} outliers between the five populations were used, both based on the Fdist method of Beaumont & Nichols (1996). Firstly, outliers were detected using the selection workbench Lositan (Antao *et al.* 2004), running 100000 simulations, specifying a confidence interval of 0.995 (CI: 0.995) and an FDR of 0.05 under the Infinite Alleles Model, using a "neutral" mean F_{ST} to remove potentially selected loci before forcing a mean F_{ST} by excluding the first set out of outliers. Secondly, F_{ST} outlier loci were identified under a Bayesian framework, as implemented in Bayescan 2.1 (Foll & Gaggiotti 2008). Bayescan was run with 5000 iterations, including a burn-in of

100 000, using a prior odds of 10, and specifying an FDR threshold of 0.05 for multiple comparison testing. As both Lositan and Bayescan use different implementations of the F_{ST} outlier method, loci were identified as being under selection when they were called as significant outliers by both programs.

Several approaches are available for identifying covarying loci of small effect in contributing to local adaptation. LFMM was used in this study, as it may be more suitable for detecting covarying loci when population structure exists (Pavey *et al.* 2015). Environmental variables included a matrix of metal-contamination data based on the six most common metals associated with the river environments for each population: arsenic (As), cadmium (Cd), copper (Cu), iron (Fe), nickel (Ni) and zinc (Zn). Data was obtained by request from the Environment Agency, and was averaged over a 15-year period (2000-2014). LFMM analysis was run in the LEA package in R (Frichot & François 2015), implementing 10000 sweeps and a burn-in of 5000 reps, as specified in Frichot *et al.* (2013). The number of latent factors was set at $K = 2$, as suggested by both the Structure analysis (see Results), and also by estimating the minimal cross-entropy criterion of the snmf function. The Gibbs sampler was run five times on the dataset in order to estimate the appropriate average LFMM parameters (z-score values). As suggested by Frichot & François (2015), p-values were adjusted using the Benjamini-Hochberg procedure (Benjamini & Hochberg 1995) and the estimation of the genomic inflation factor (λ). An FDR of 0.5 and adjusted p-values <0.001 were used to test for significance. Estimated FDR, λ and histograms of adjusted p-values for each metal can be found in Table S1 and Figure S1.

Because the spatial structure of the five populations has the potential to mask detection of outlier loci relating specifically to the two metal-impacted populations, we conducted additional DAPC analyses in order to assess the effects of these outlier loci on distinguishing population structure. Two DAPC plots were produced, the first using F_{ST} outlier loci, and the second using loci identified by LFMM. The number of PCs that described >90% variance in the dataset were used, and the first 3 discriminant functions were retained for analysis. Based on the results of this analysis, and others (see Results), additional tests were conducted to detect Hayle and Red River outlier loci. Both Lositan and Bayescan were run with the same parameters as specified above,

except that we specified the detection of F_{ST} outliers between the (i) Hayle vs. Clean groups and (ii) the Red River vs. Clean groups.

Exploring the architecture of divergent loci between clean and metal ecotypes

To assess the genetic makeup of loci under selection and how these loci are diverged between the population groups, we assessed the haplotype gene diversity (H_E) and divergence based on *Phi*$_{ST}$ (Φ_{ST}) for each outlier locus. Diversity indices were compared across the populations in order to assess the relative role of homozygous fixation or heterozygous advantage in driving metal-tolerance. This also helped discern whether selection was based on standing genetic variation or *de novo* mutation. Differentiation of loci between the populations were assessed in order to decipher the relative contribution these loci have on driving differential allele frequencies in the metal-impacted and clean-river populations.

Gene ontology of loci under divergent selection

All loci identified as outliers (comparison of all populations; Hayle vs. Clean; Red River vs. Clean) were assigned gene ontology information using megablast against the RefSeq Atlantic salmon genome (ICSASG v2; NCBI Accession number: GCF_000233375.1). Ontology was assigned based on the top hit, using an e-value of 1e-34 for significance. For loci that did not fall within a coding region, ontology was assigned based on the closest genes on either side of the alignment (3' or 5'), assuming that the RAD locus would be under selection due to linkage.

We also performed Gene Ontology (GO) analysis on the outliers identified by the F_{ST} outlier methods for the Hayle and the Red River trout populations. In a conservative approach, ontology was assigned as outlined above, excluding any loci that aligned to multiple chromosomes, or to multiple places on the same chromosome. Gene lists were converted to Uniprot official gene symbols based on ontology to *Homo sapiens*, and uploaded to The Database for Annotation, Visualization and Integrated Discovery (DAVID; v. 6.8; Huang *et al.* 2008, 2009). We used *Homo sapiens* as the background for the gene lists for exploring GO term enrichment, and explored GO Biological Processes (GO BP), GO Cellular Component (GO CC) and GO Molecular Function (GO MF).

Genome-wide signals of local adaptation generated by alignment to a 'non' reference genome

In order to identify genomic islands of divergence driving metal-tolerance in the two metal-impacted populations, we assessed genome-wide signals by producing a 'large hadron collider' (LHC) plot using the python script lhc_plot.py in Stacks (v.1.46) (Catchen *et al.* unpublished). Given that the brown trout does not currently have a reference genome, we used information obtained by alignment of trout RAD loci to the Atlantic salmon (*Salmo salar*) genome ICSASG v2 (NCBI GCA_00023375.4). To do this, catalog loci built *de novo* in Stacks were aligned using gsnap (Wu & Nacu 2010), specifying a maximum of 5 mismatches (-m 5), preventing indel alignment (-i 2) and turning off terminal alignments (--min-coverage=0.95) to prevent reads from being soft-masked. We integrated unique alignments (unpaired_uniq) from the alignment files back into the *de novo* dataset using the integrate_alignments.py script in Stacks (Paris *et al.* 2017). Because considerably fewer loci were detected using the integrate method (see Results), we filtered the loci in `populations` differently to the *de novo* loci, this time specifying a locus be present in 50% of individuals (-r 0.5), at a minor allele frequency of 10% (--min_maf 0.1) and a maximum observed heterozygosity of 70% (--max_obs_het 0.7).

For the LHC plots, we plotted the smoothed *Phi*$_{ST}$ (Φ_{ST}) of (i) the Hayle vs. Clean, and (ii) the Red River vs. Clean. Smoothed Φ_{ST} was calculated using a Gaussian weighting function sliding window implemented in Stacks (Catchen *et al.* 2013), using a sigma (σ) size of 500kb. To assess whether regions under divergence were correlated with reduced haplotype diversity, we plotted the raw haplotype diversity of the Hayle and Red River trout populations next to the respective regions of the genome under genomic divergence.

RESULTS

Quality control, filtering and RAD loci construction

A total of ~218 million reads were retained after demultiplexing and cleaning the raw sequence data, averaging 2.2 million reads (± 186,215 reads) per individual (min. 658,336 – max. 14,845,672). An individual from the Carrick Roads sample was removed due to low coverage, resulting in an analysis of a total of 99 individuals across the five putative populations. The total number of loci built

across the first catalog (comprising all individuals) was 271,222 of which 104,806 were polymorphic, and contained a total of 315,612 SNPs. The improved catalog, comprised of only the highest coverage individuals (n=75), consisted of 269,146 loci, 54,141 polymorphic loci and 160,041 SNPs, representing an average of ~3 SNPs per RAD locus. Specifying stringent criteria for locus processing (r80%, MAF 10%, maximum heterozygosity 70%) resulted in a reliable set of 5,825 polymorphic loci, which were used for all downstream *de novo* analyses. Across these polymorphic loci, 4,660 contained one SNP, 820 contained two SNPs, 212 contained three SNPs, 92 contained four SNPs and 41 contained five SNPs, giving a total of 7,509 SNPs.

Alignment of *de novo* loci to the Atlantic salmon genome using the integrated method showed that 107,910 aligned uniquely, which represented approximately 40% of all *de novo* assembled loci. Filtering loci for investigating islands of divergence between the metal-impacted populations resulted in 25,595 loci (7,907 polymorphic) for the Hayle vs. Clean comparison and 27,547 loci (8,382 polymorphic) for the Red River vs. Clean comparison.

Population genetics statistics

Mean expected heterozygosity (H_E) ranged from 0.31-0.34 and observed heterozygosity (H_O) ranged from 0.24-0.31 (Table 2). Metal-impacted populations showed the lowest H_O: Hayle (0.24), Red River (0.28), whereas the clean populations all showed H_O>0.3. Values of the inbreeding coefficient (F_{IS}) suggested that the metal-impacted populations contained an excess of homozygotes relative to Hardy-Weinberg equilibrium (Hayle F_{IS} 0.24; Red River F_{IS} 0.17), whereas the clean populations showed lower F_{IS} values (Table 2). Estimates of effective population size (N_e) showed that metal-impacted trout populations demonstrated considerably lower N_e compared to populations from neighbouring clean rivers; the Red River had the lowest N_e, followed by the Hayle (Table 2).

Table 2. Population genetics statistics for each population: H_E (expected heterozygosity); H_O (observed heterozygosity) and F_{IS} (Wright's inbreeding coefficient) with 95% confidence intervals and N_e (effective population size) with 95% confidence intervals for the three clean populations: Camel; Carrick Roads and Gannel (denoted by *) and the two metal-impacted populations: Hayle and Red River (denoted by †).

Population	H_E	H_O	F_{IS} (95% CIs)	N_e (95% CIs)
Camel *	0.34	0.31	0.09 (0.080-0.095)	490.2 (454.9-531.3)
Carrick Roads *	0.33	0.30	0.11 (0.099-0.115)	125.2 (122.4-128.0)
Gannel *	0.33	0.30	0.10 (0.087-0.102)	645.7 (586.2-718.7)
Hayle †	0.31	0.24	0.24 (0.230-0.246)	69 (68.1-69.9)
Red River †	0.33	0.28	0.17 (0.162-0.178)	57.2 (56.5-57.9)

Table 3. Pairwise F_{ST} calculated between each pair of populations. The three clean populations: Camel; Carrick Roads and Gannel are denoted by * and the two metal-impacted populations: Hayle and Red River are denoted by †. F_{ST} values are presented above the diagonal, respective significance values are shown below the diagonal. p<0.001 is represented as ***

	Camel *	*Carrick Roads* *	*Gannel* *	*Hayle* †	*Red River* †
Camel *		0.029	0.039	0.112	0.055
Carrick Roads *	***		0.058	0.125	0.069
Gannel *	***	***		0.112	0.067
Hayle	***	***	***		0.094
Red River †	***	***	***	***	

The mean global F_{ST} across all populations was 0.062 (p=0.001). All pairwise-F_{ST} comparisons were significant (p=0.001; Table 3), with the highest F_{ST} values occurring between the Hayle and the three clean-river populations (F_{ST} 0.112-0.125), followed by Hayle vs. Red River (F_{ST} 0.094), and then Red River vs. the three clean-river populations (F_{ST} 0.067-0.069), with the exception of Camel vs. Red River (F_{ST} 0.055). The lowest F_{ST} values were observed between the Camel and the Carrick Roads (F_{ST} 0.029) and the Camel and the Gannel (F_{ST} 0.039). Pairwise F_{ST} values obtained in this study were higher compared to results obtained previously using neutral microsatellite markers (Paris *et al.* 2015; Table S2).

Together with the high pairwise F_{ST} observed between the Hayle and the Red River (F_{ST} 0.094, p<0.001), significant divergence between these populations was also evidenced by the AMOVA results. The AMOVA comparing the nesting of the three clean populations and the two metal populations, showed a higher proportion of variance was explained when the two metal populations were specified as separate groups (Hayle and Red River; F_{CT} = 7.9%, F=0.079, p<0.05), compared to when they were grouped together as one 'metal' group (F_{CT} = 4.4%, F=0.044, p=0.102).

F_{ST} differentiation was not associated with geographic distance, or isolation-by-distance (IBD); r = 0.104, p=0.228 (Figure 1A). The Mantel test exploring the relationship between genetic distance and the overall levels of metal pollution for each river was not significant (r = 0.013, p=0.557). However, effects of particular metal contaminants in driving genetic differentiation were statistically significant for cadmium (r = 0.332, p=0.05), nickel (r = 0.762, p=0.029) and zinc (r = 0.649, p=0.034), but non-significant for arsenic (r = 0.007, p=0.77), copper (r = 0.17, p=0.113) and iron (r = 0.038, p=0.355) (Figure 2).

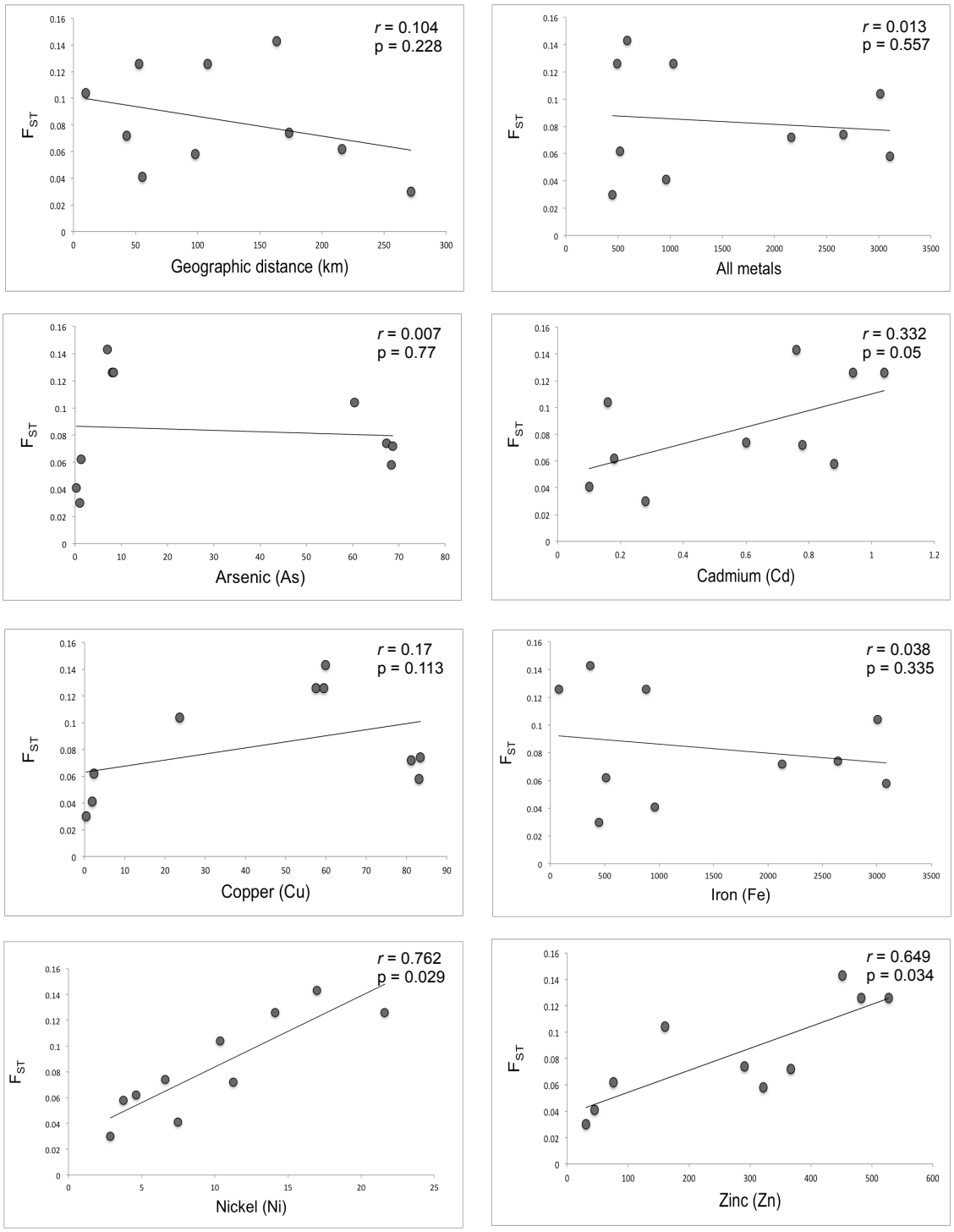


Figure 2. Mantel tests of isolation-by-distance (IBD) and isolation-by-environment (IBE). (A) IBD mantel test of pairwise genetic differentiation (F_{ST}) and geographic distance (km). (B) IBE mantel test of F_{ST} and the *Euclidean* distances calculated for all metal contaminants, (C) IBE F_{ST} and arsenic (As); (D) IBE F_{ST} and cadmium (Cd); (E) IBE F_{ST} and copper (Cu); (F) IBE F_{ST} and iron (Fe); (G) IBE F_{ST} and nickel (Ni) (H) IBE F_{ST} and zinc (Zn).

Population structure

The program Structure identified two genetic clusters (*K*=2) based on the genotype frequencies of the 5,825 polymorphic loci across all five putative populations (Figure 3A). This showed that the Hayle trout (metal-impacted), formed a distinct group and that the three clean-river groups had similar genetic profiles. The Red River trout (metal-impacted) appeared to comprise of an admixture of genotypes representative of both the Hayle trout and the clean-river genotypes.

Results of the DAPC separated all groups along the three principal components (Figure 3B). The first principal component (eigenvalue = 41674) separated the Hayle trout from all other groups. The second principal component represented less variability (eigenvalue = 22105), and separated the Gannel and Carrick Roads trout from the Red River, Camel and Hayle samples. The third axes represented significantly less variance (eigenvalue = 7593) and did not distinguish any further groupings of the samples.

The neighbour-joining dendrogram of Nei's genetic distances showed high bootstrap values of 100 for all branches (Figure 3C). This again showed that the two metal-impacted trout populations are genetically distinct from other populations, and each other, appearing on separate branches of the dendrogram, with the Hayle population showing the longest branch length. Amongst the clean populations, the Camel and the Carrick Roads appeared to be more similar to one another, and the Gannel population formed a separate branch.

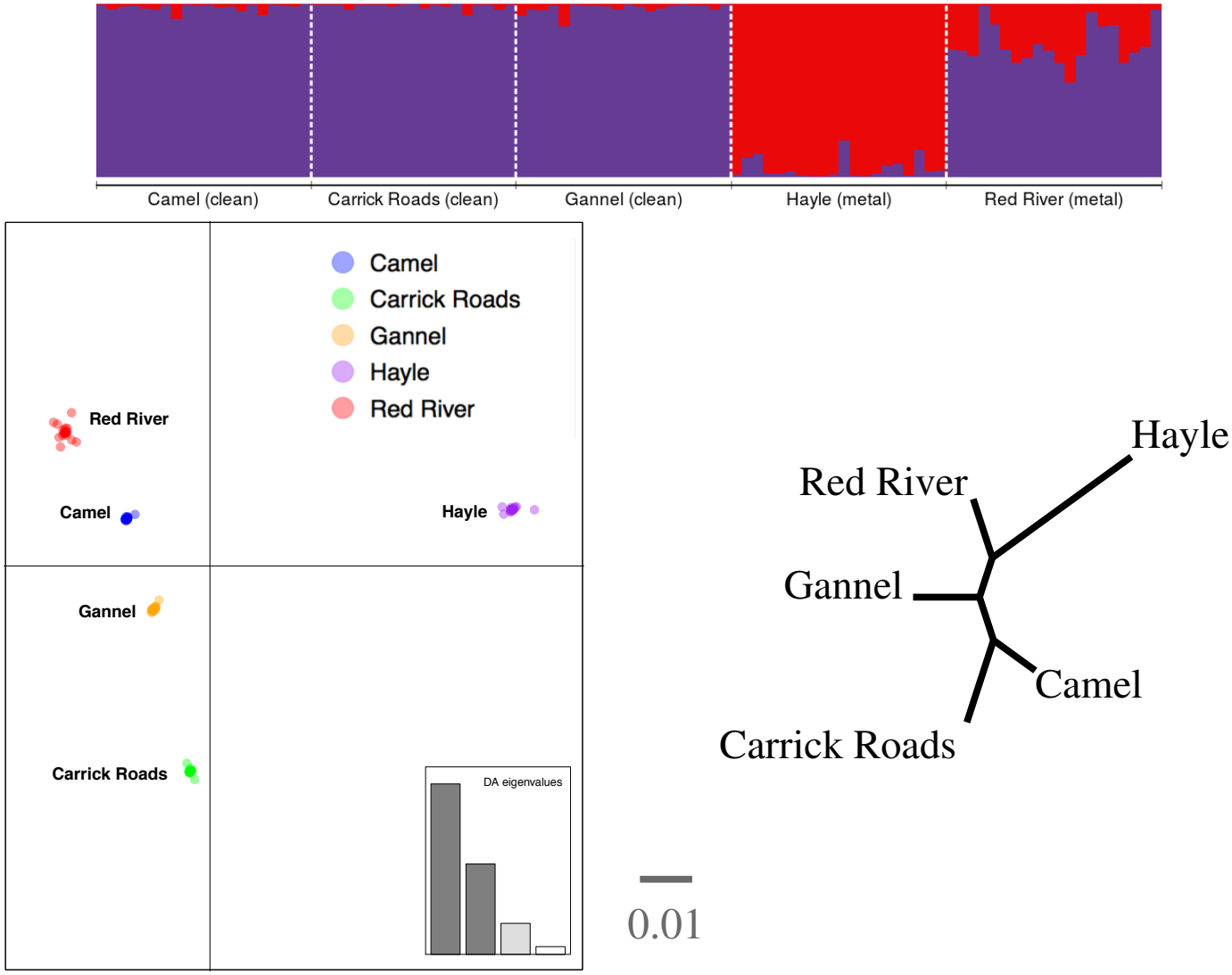


Figure 3. Population structure analyses based on 5,825 polymorphic loci for the five trout populations. (Details continue on next page).

Figure 3. Population structure analyses based on 5,825 polymorphic loci for the five trout populations: Camel (clean); Carrick Roads (clean); Gannel (clean); Hayle (metal) and Red River (metal). (A) Structure plot showing the most likely number of clusters (*K*=2). (B) DAPC ordination plot of the five populations. Clusters of coloured circles represent individuals within each population. Discriminant analysis (DA) eigenvalues are shown in the bottom right, showing the relative proportion of each eigenvalue in separating the population clusters. (C) Unrooted neighbour joining tree calculated using a matrix of genetic distances based on Nei's genetic distance.

Analysis of loci under selection in metal-impacted populations

Analysis of F_{ST} outlier loci comparing all five populations identified 42 outlier loci (0.7% of all loci) using Lositan and 33 outlier loci (0.6%) using Bayescan (Table S3). Combining the F_{ST} outliers from both methods identified 12 common outlier loci (0.2%). DAPC of these divergent loci showed a clear distinction between the two metal-impacted populations (Hayle and Red River) compared to the three clean-river populations, which all clustered together (Figure 4A). The first principal component (PC) had a high eigenvalue (eigenvalue = 380) and separated the Hayle and a large proportion of the Red River apart from the clean populations. The second PC comprised an eigenvalue of just 54, separating the Hayle trout from the Red River trout and the third PC represented an eigenvalue of 7 and did not distinguish the metal-impacted populations any further.

Considerably more loci were identified as being under selection when additional F_{ST} outlier analyses we performed to explore patterns specifically between the Hayle vs. Clean and Red River vs. Clean. Hayle-specific selection tests revealed seven Bayescan outliers and 155 Lositan outliers, with all seven of the Bayescan outliers also identified by Lositan. The Red River-specific selection tests identified 11 Bayescan outliers and 74 Lositan outliers, with all 11 Bayescan outliers identified by Lositan. Only one locus (11539) was identified as under selection when comparing all populations, Hayle-specific and Red River-specific outliers. Another locus (18205) was identified as under selection when comparing all populations, and also specifically in the Red River, whereas three loci (38481, 42147, 46160) were identified as under selection in the comparison of all populations, and also in the Hayle-specific analysis. This suggests that the metal-population-specific F_{ST} outlier tests were more sensitive in unpicking divergent loci specific to each metal-impacted population, and potentially loci associated with selection to unique cocktails of metal contaminants.

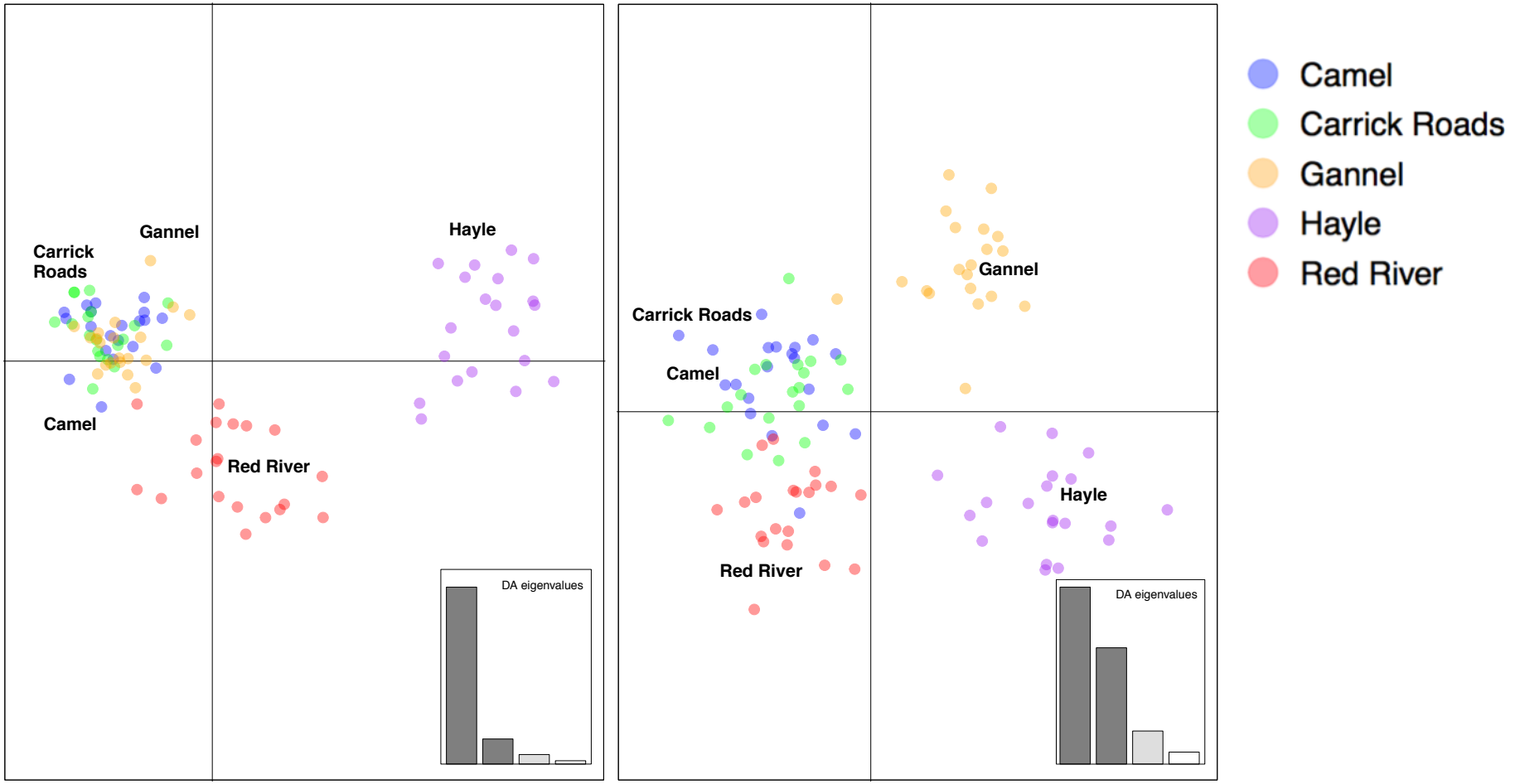


Figure 4. Discriminant analysis of principal components (DAPC) using (A) F_{ST} outlier loci (B) LFMM outlier loci for the five trout populations: Camel (blue); Carrick Roads (green); Gannel (orange); Hayle (purple) and Red River (red).

The LFMM analysis identified a total of 29 outlier loci associated with the metal contamination data (0.5% of all loci), 24 identified in common across all metals, one locus correlated with arsenic, nickel and zinc, and four loci associated uniquely with arsenic. One locus was identified as under selection in both LFMM and Bayescan (10603), and one locus was identified in both LFMM and Lositan (11460), but no loci were found to be in common across all three outlier methods. The DAPC of the LFMM outliers showed a different pattern of population structure compared to the F_{ST} outlier loci (Figure 4B). The first two principal components (PCs) captured the large majority of the variance (first eigenvalue = 163; second eigenvalue = 107), whilst the third discriminated very little of the variance in comparison (eigenvalue = 30). Interestingly, the first two PCs separated the known metal-impacted populations (Hayle and Red River) whilst also considerably separating the Gannel trout away from the clean samples. Indeed, the Gannel group was significantly more distinct even when compared to the Red River group, distinguishing as a separate cluster along the first PC.

Functional gene ontology and genetic architecture of loci under divergent selection

For the F_{ST} outlier analysis comparing all populations, all twelve loci showed significant alignment to the salmon genome, ten of which were positioned within protein coding regions (Table 4; Table S4) and two of which fell within 24 to 158 Kb of nearby coding regions. Amongst the annotations were genes involved in protein coating and trafficking, DNA damage response and transcriptional regulation.

Twenty-two of the 24 outlier loci identified by LFMM as being common to all metals showed significant hits to the salmon genome (Table S4), thirteen fell within coding regions, the remaining nine loci were within approximately 5 to 656 Kb from coding regions. Annotations included several genes associated with functions involved in metal-ion and calcium-ion binding, as well as protein ubiquitination and immune response. The four loci that were associated with arsenic all fell within coding regions. The locus identified in association with arsenic, nickel and zinc showed homology to RNA-binding protein 4 (RBM4).

Overall, no general trend was observed for either increased or decreased gene diversity (H_E) across the various groups of F_{ST} outlier loci (Table 4). However, some loci exhibited increased diversity within the two metal-impacted populations; for example, locus 11539 (identified as an outlier in both the Hayle and the Red River) was annotated as being similar to antizyme inhibitor 1 (AZIN1). This locus was variable across all populations, but showed higher diversity in the metal-impacted populations (Hayle H_E 0.203; Red River H_E 0.383), compared to the clean populations (Clean H_E 0.179). As expected, differentiation at this locus was high between the Hayle vs. Clean (Φ_{ST} 0.77) and the Red River vs. Clean (Φ_{ST} 0.65), and very low between the Hayle vs. Red River (Φ_{ST} 0.02). Locus 46160, which aligned to an intergenic region, showed complete homogeneity across all of the clean populations (H_E 0), whereas both metal populations showed significantly increased diversity (Hayle H_E 0.43; Red River H_E 0.476). On the other hand, other loci showed homozygosity (suggestive of fixation) in metal-populations, where increased diversity was evidenced in the clean populations. For example locus 63058, which aligned to an intergenic region, showed high diversity (H_E 0.47) across the clean-river populations, but was fixed in both the Hayle (H_E 0) and the Red River (H_E 0).

Seven of the 10 Red River-specific outliers showed alignment to the salmon genome, of which three fell within coding regions, and the remaining four were within 4 to 6381 Kb of coding regions (Table 4; Table S4). Locus 18205 (annotated as showing homology to lamina-associated polypeptide 2; LAP2-like) was identified as a significant outlier in the analysis of all populations, and also in the Red River-specific analysis. This locus showed low diversity in the clean-river populations (H_E 0.122), considerably higher diversity in both metal-impacted populations (Hayle H_E 0.386; Red River H_E 0.451) and showed high differentiation between the Red River vs. Clean (Φ_{ST} 0.65), low differentiation between the Hayle vs. Clean (Φ_{ST} 0.14), and higher differentiation when comparing the Red River and Hayle populations (Φ_{ST} 0.29), supporting that this locus is a Red River-specific outlier.

Table 4. Details of each F_{ST} outlier RAD locus including number of SNPs, gene diversity (H_E) and genetic differentiation Phi_{ST} (Φ_{ST}) between the groups (Clean, the Hayle, and the Red River) of populations.

				Gene Diversity (H_E)			Phi_{ST} (Φ_{ST})		
Comparison	*Locus ID*	*Coding region (or distance to nearest coding region)*	*No. SNPs*	*Clean*	*Hayle*	*Red River*	*Clean vs. Hayle*	*Clean vs. Red River*	*Hayle vs. Red River*
All Clean v All Metal	3908	clathrin heavy chain 1 (CLTC)	3	0.234	0.475	0.487	0.46	0.44	-0.03
	4204	coiled-coil domain-containing protein 73 (CCDC7)	1	0.204	0.358	0.491	0.64	0.21	0.24
	11539	antizyme inhibitor 1 (AZIN1)	1	0.179	0.203	0.383	0.77	0.65	0.02
	18205	lamina-associated polypeptide 2-like (LAP2)	1	0.122	0.386	0.451	0.14	0.65	0.29
	35593	pro-neuregulin-3, membrane-bound iosform-like (NRG3)	1	0.364	0.226	0.489	0.55	0.27	0.14
	38481	ras-like protein family member 10B (RASL10B)	1	0.114	0.224	0.05	0.82	0	0.84
	42147	158356 bp at 5' side: zinc finger protein 585A-like (ZNF585A) / 64152 bp at 3' side: EMILIN-1-like (EMILIN1)	1	0.352	0.332	0.603	0.65	0.05	0.46
	43057	disintegrin and metalloproteinase domain-containing protein 8-like (ADAM8)	1	0.383	0.371	0.471	0.41	0.01	0.27
	46160	68406 bp at 5' side: uridine phosphorylase 2-like (UPP2) / 24953 bp at 3' side: plakophilin-4-like (PKP4)	1	0	0.431	0.475	0.82	0.45	0.24
	46300	aminoacyl tRNA synthetase complex-interacting multifunctional protein 1 (AIMP1)	1	0.413	0.203	0.512	0.5	0.1	0.25
	56666	ankyrin repeat domain-containing protein 32 (ANKRD32)	1	0.504	0.203	0.224	0.27	0.25	-0.03
	59777	galanin receptor type 1-like (GALR1)	1	0.391	0.309	0.157	0.45	0.07	0.69
Hayle v Clean	11539	antizyme inhibitor 1 (AZIN1)	1	0.179	0.203	0.383	0.77	0.65	0.02

	21668	46118 bp at 5' side: phosphatidylinositol 3,4,5-trisphosphate-dependent Rac exchange factor (PREX1) / 63226 bp at 3' side: RING finger protein 223-like (RNF223)	1	0.5	0	0.494	0.48	0.1	0.28
	38481	ras-like protein family member 10B (RASL10B)	1	0.114	0.224	0.05	0.82	0	0.84
	42147	--	1	0.352	0.332	0.603	0.65	0.05	0.46
	46160	68406 bp at 5' side: uridine phosphorylase 2-like (UPP2) / 24953 bp at 3' side: plakophilin-4-like (PKP4)	1	0	0.431	0.475	0.82	0.45	0.24
	55120	F-box/LRR-repeat protein 17 (FBXL17)	1	0.27	0.294	0.649	0.66	0.14	0.35
	71248	ephrin type-A receptor 7-like (EPHA7)	1	0.326	0.383	0.401	0.37	0.11	0.39
Red River v Clean	11539	antizyme inhibitor 1 (AZIN1)	1	0.179	0.203	0.383	0.77	0.65	0.02
	14054	--	1	0	0.385	0.514	0.38	0.67	0.1
	14577	4013 bp at 5' side: dnaJ homolog subfamily B member 5 (DNAJB5) / 5157 bp at 3' side: transitional endoplasmic reticulum ATPase (VCP)	1	0.489	0.322	0	0.08	0.31	0.18
	18205	lamina-associated polypeptide 2 (LAP2)	1	0.122	0.386	0.451	0.14	0.65	0.29
	23716	rho GTPase-activating protein 32 (ARHGAP32)	2	0.573	0.522	0.157	0.12	0.35	0.15
	25067	SUN domain-containing ossification factor-like isoform X2	1	0.11	0.501	0.5	0.38	0.57	0.02
	27901	guanine nucleotide-binding protein G(i) subunit alpha-1-like (GNAI1)	1	0.521	0.486	0.315	0.09	0.32	0.09
	29115	--	1	0.513	0.273	0	0.17	0.35	0.14
	32833	104486 bp at 5' side: receptor-type tyrosine-protein phosphatase F-like (PTPRF) / 381428 bp at 3' side: uncharacterized protein LOC106584099	3	0.73	0.409	0.376	0.25	0.3	-0.01
	60772	104114 bp at 5' side: zinc finger E-box-binding homeobox 2-like / 137885 bp at 3' side: kynureninase-like (KYNU)	1	0.13	0.385	0.491	0.13	0.57	0.21
	63058	40115 bp at 5' side: fidgetin-like (FIGN) / 43354 bp at 3' side: complement C1q-like protein 2 (C1QL2)	1	0.47	0	0	0.27	0.28	0

Five of the six Hayle-specific outlier loci showed alignment to the salmon genome, three of which were assigned ontology to coding regions, two in intergenic regions within 25 to 68 Kb of coding regions, and one could not be assigned any gene ontology (Table 4; Table S4). Loci 38481, annotated as ras-like protein family member 10B, showed high diversity in the Hayle trout (HE 0.224), average diversity in the clean-river populations (HE 0.114) and yet considerably lower diversity in the Red River population (H_E 0.05). Correspondingly, this locus showed extremely high differentiation between Hayle vs. Clean (Φ_{ST} 0.82) and also Hayle vs. Red River (Φ_{ST} 0.84), reinforcing the role for this locus in local adaptation of Hayle trout to the river-specific environment. Moreover, locus 42147, for which no gene ontology could be assigned, showed similar levels of diversity in the clean (H_E 0.352) and Red River populations (H_E 0.332), but considerably higher diversity in Hayle trout (H_E 0.603). These differences in diversity were also correlated with high differentiation between the Clean vs. Hayle (Φ_{ST} 0.65) and Red River vs. Hayle (Φ_{ST} 0.46). Finally, locus 46160 lying in an intergenic region 25 to 68 Kb between two protein coding regions (uridine phosphorylase 2-like and plakophilin-4-like), showed no diversity across any of the clean-river populations (H_E 0), yet high diversity in the two metal-impacted populations (Hayle H_E 0.431; Red River H_E 0.475). These differences in diversity were matched by strong patterns of differentiation between the metal-impacted and clean-river populations (Hayle Φ_{ST} 0.819, Red River Φ_{ST} 0.448) and lower, but still noticeable, differentiation between the Hayle and Red River populations (Φ_{ST} 0.239).

Islands of genomic divergence

The large hadron collider plot (Figure 5) showed that the Hayle population exhibited higher genomic divergence compared to the Red River population, but no specific chromosome(s) alone were identified as showing strong underlying patterns of divergence between clean and metal-impacted populations. Several regions of the genome showed patterns of genomic divergence that were mirrored in both metal-impacted populations (Hayle and Red River) e.g. ssa02, ssa11, ssa21. Several genomic regions showed river-specific patterns of differentiation. For example, Hayle-specific regions can be seen on ssa04, ssa06 and ssa15 and Red River-specific regions can be seen on ssa14, ssa16 and ssa17. Intriguingly, different regions on the same chromosome showed differential divergence between the Hayle and the Red River populations (e.g. ssa01, ssa07, ssa28). In some cases, regions with increased Φ_{ST} also showed a correlated reduction in haplotype diversity (e.g. ssa25, ssa27), but generally increased Φ_{ST} did not correlate with reduced haplotype diversity (e.g. ssa01, ssa20). Overall, these patterns suggest that selection is stronger in the Hayle compared to the Red River, yet is neither specific nor shared between metal-impacted trout from the Hayle and Red River. Furthermore, selection does not appear to be occurring as a hard sweep on localised regions of the genome (see Discussion).

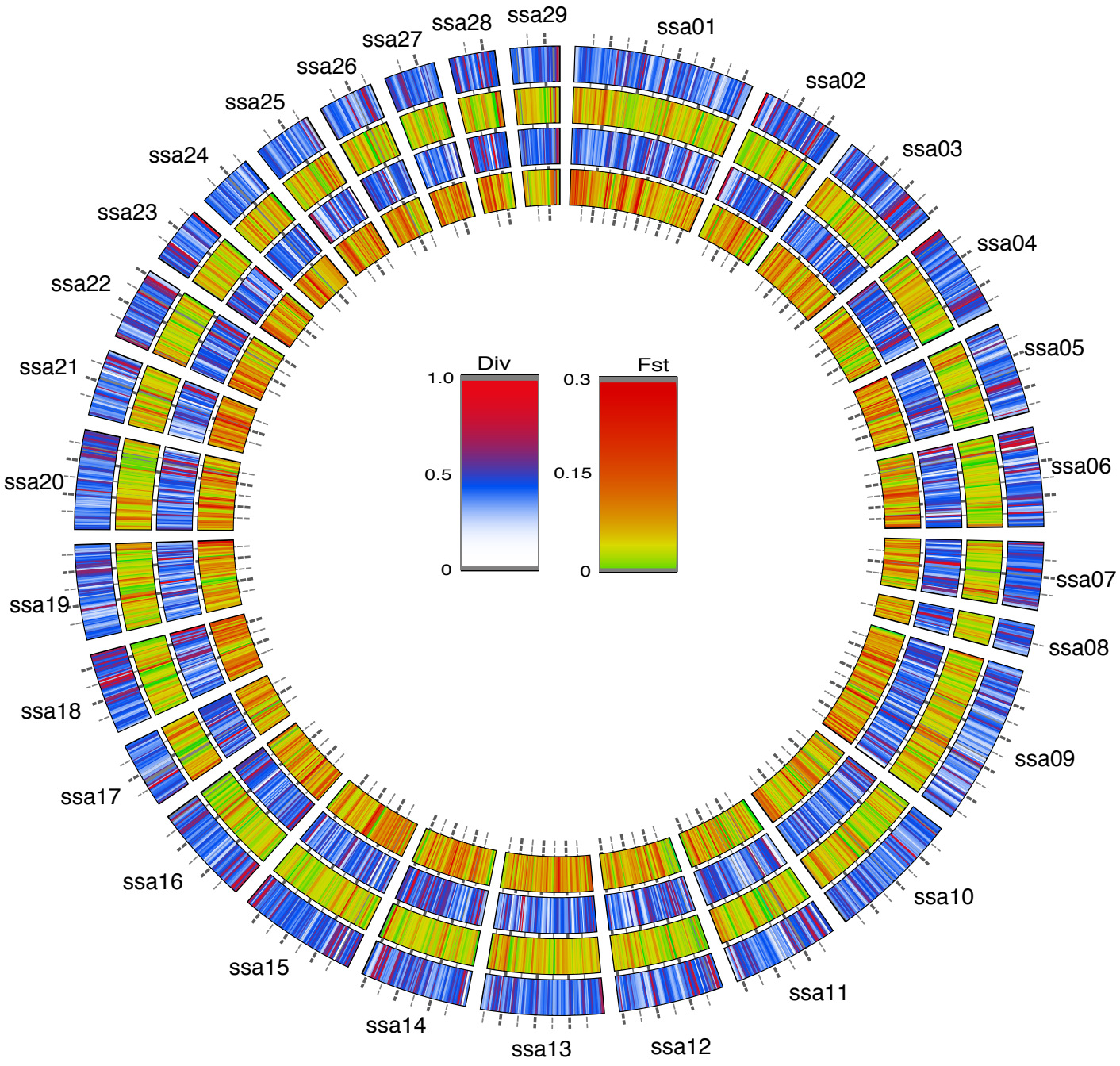


Figure 5. Islands of genomic divergence and haplotype diversity between the two metal-impacted populations compared to the clean-water populations. The two inner circles represent smoothed F_{ST} and haplotype diversity of trout in the Hayle vs. Clean. The two outer circles represent smoothed F_{ST} and haplotype diversity of trout in the Red River vs. Clean. The gradients of colours represent average diversity (Div.) or divergence (F_{ST}), with haplotype diversity represented from 0 to 1 and F_{ST} divergence represented from 0 to 0.3.

DISCUSSION

Pressures from anthropogenic activity are increasingly recognised as key drivers in the evolutionary trajectory of populations. Brown trout populations occupying rivers affected by historical mining activity, and polluted by a patchwork of toxic metal pollution, offers a unique opportunity to assess the role of local adaptation to human-altered environments. The results illustrate that metal-impacted populations show distinct genetic patterns, including decreased heterozygosity and low effective population sizes. Further, the two metal-impacted populations, separated by just a few kilometres showed genetic divergence compared to the three clean-river populations, which are separated by hundreds of kilometres. Moreover, each metal-impacted population demonstrated a unique set of outlier loci, suggesting the role of different selective traits in conferring local adaptation to unique cocktails of metal pollutants in populations of metal-tolerant trout.

Metal-impacted populations demonstrate distinct patterns of genetic diversity and genetic differentiation

The effects of human-impacted environments in driving patterns of altered genetic diversity and reductions in population size in natural populations are common phenomena (Stockwell *et al.* 2003; Smith & Bernatchez 2008; Sgrò *et al.* 2011). In comparison to trout from clean rivers, metal-impacted trout exhibited reduced heterozygosity and low effective population sizes. This suggests that modified environmental conditions arising as a result of historical mining activity have significantly altered the genetic makeup of these populations. These results mirror the genetic patterns exhibited previously by microsatellite markers, which demonstrated reduced genetic diversity and evidence of population bottlenecks in mining-impacted trout populations (Paris *et al.* 2015). In particular, chemical pollutants have been shown to result in detrimental population-level effects in fish populations (Bickham *et al.* 2000; Hamilton *et al.* 2016). For example, in threespine stickleback (*Gasterosteus aculeatus*) inhabiting polluted rivers across England (Santos *et al.* 2013) and in yellow perch (*Perca flavescens*) occupying metal-impacted lakes in Canada (Bourret *et al.* 2008).

Such genetic signatures may also be indicative of selection (Charlesworth 1996; Gillespie 1997; Andolfatto 2001), especially considering that selection has a higher chance of success in populations that have experienced a bottleneck if it acts on sites already segregating in the population (Hermisson & Pennings 2005). Depending on the strength of the selection pressure, the rates of recombination, and the degree of linkage disequilibrium (Smukowski & Noor 2011; Cutter & Payseur 2013; Haasl & Payseur 2016), selection can significantly reduce genetic variability of neutral regions associated with selective advantage (Maynard-Smith & Haigh 1974; Barton 2000). Concurrently, processes of extensive purifying selection may also reduce heterozygosity at sites linked to selection (Gordo *et al.* 2002; van Oosterhout *et al.* 2006; Charlesworth 2012). For example, reduced diversity was evidenced in fourfold degenerate (silent) sites in the collared flycatcher (*Ficedula albicollis*), likely through processes of purifying selection operating specifically on these regions, as well as through hitchhiking via linkage to nearby nonsynonymous sites (Dutoit *et al.* 2016). In a survey of 12 QTLs distributed across the dog genome likely responsible for artificial selection in breed-specific traits, elevated and extended levels of haplotype homozygosity were shown to occur around each sweep locus, which also increased with the frequency of each selected allele (Schlamp *et al.* 2016). However, such genetic patterns are not always evidenced. Despite a long period of sustained selection pressure in populations of *Drosophila magna*, no reductions in genetic diversity or effective population size were detected (Orsini *et al.* 2016) and patterns of reduced diversity at toll-like receptor (TLR) loci in Berthelot's pipit (*Anthus berthelotii*) were more strongly affected by demography and genetic drift rather than selective processes (Gonzalez-Quevedo *et al.* 2015).

The contribution of a mosaic of neutral and non-neutral genetic forces in driving distinct genetic differences in metal-impacted populations is also compelling when we consider the observed patterns of population structure. Strong patterns of population structure are typically evidenced in brown trout populations; in an assessment of 82 populations of the species across 29 rivers in southwest England, eighteen microsatellite markers differentiated substantially high levels of genetic structuring across the region, and in particular, these patterns were strongly associated with geography (King *et al.*

2016). Even within-catchments, brown trout populations can show a high degree of intra-river population structuring and isolation-by-distance (Carlsson *et al.* 1999; Griffiths *et al.* 2009; Kanno *et al.* 2011; Paris *et al.* 2015). In comparison, such strong patterns of population structure are not so apparent in Atlantic salmon (Griffiths *et al.* 2010; Ellis *et al.* 2011). This suggests that, in brown trout, even in the absence of strong selection pressures imposed by metal pollution, there is also a marked population structure driven by other genetic and geographic processes.

In addition, the stochastic phylogeography of the brown trout species (Bernatchez 2001; McKeown *et al.* 2010) also imposes an underlying genetic structure that must be considered. Likewise, both historical and contemporary processes were shown to have contributed to the diversity and distribution of the galaxid fish, *Galaxias platei,* in Patagonia (Vera-Escalona *et al.* 2015) and patterns of genetic diversity in metapopulations of longnose suckers (*Catostomus catostomus*) in Canada, were the result of colonisation from a palaeolake (Salisbury *et al.* 2016). Over time, these phylogeographic signatures are likely to be further enforced via selection operating in each population to the local ecological features of each river (e.g. Meier *et al.* 2011; Carlson & Nilsson 2001). Such forces would further strengthen population structure through reduced gene flow between catchments, resulting in smaller local effective population sizes (Hansen *et al.* 2002; Fraser *et al.* 2007). Lastly, natal homing is known to be an important biological mechanism in promoting local adaptation in salmonids (Taylor 1991; Fraser *et al.* 2011), and would further augment the above processes (Carlsson & Nilsson 2000). Together, these factors suggest that both historical and contemporary genetic processes have contributed to the patterns observed; such mechanisms may also enforce the power of selection to allow metal-impacted trout populations to locally adapt to the metal-rich conditions of their native rivers.

The observed patterns of population structure are also indicative of a significant contribution of genetic drift in driving the observed differences between metal-impacted populations. Although it is challenging to tease apart the relative contributions of selection and genetic drift, it can be predicted that the patterns observed likely represent the interplay between both genetic forces. For example, in island populations of mice separated by only 5 to 10km, different

populations contained very different numbers of chromosomes (Britton-Davidian *et al.* 2000). As these populations showed low effective population sizes, this was likely caused by genetic drift; however, the maintenance of different chromosomal variation also suggests that selection was favouring the most common chromosomal form in each population (Britton-Davidian *et al.* 2000). Finally, the potency of drift and selection acting on the metal-impacted populations may be emphasised over time given that movement of clean-river trout into metal-impacted rivers may be restricted by metal pollution acting as a chemical barrier to migration (Durrant *et al.* 2011; Paris *et al.* 2015; Tétard *et al.* 2016; Gray *et al.* 2016). The strength of these myriad processes in shaping the genetic makeup of these populations should be kept in mind, and inarguably act in concert in producing patterns of genetic diversity and structure. However, despite this, we still observed compelling evidence for local adaptation associated specifically with populations residing in metal-impacted rivers.

Evidence of population-specific metal-driven local adaptation

In all assessments of genetic differentiation, it was apparent that both the Hayle and the Red River trout showed divergence from the clean-river populations, and from one another. No patterns of isolation-by-distance were evidenced amongst the five populations and this was likely a result of the high F_{ST} values between the two metal-impacted populations, and low F_{ST} values between the geographically distant clean populations. This is especially striking given that the mouths of the River Hayle and Red River are separated by ~9 km yet the trout populations showed an F_{ST} approaching 0.1, whereas, for example, the Camel and the Carrick Roads are separated by ~270 km, yet showed a smaller F_{ST} value of just 0.029. Geographic distance is not the primary factor driving the high genetic differentiation between metal-impacted populations.

If two populations are geographically close, but experience substantially different selective environments, the expected differentiation due to non-selective factors is small, whilst selection could still be extremely effective (Lotterhos *et al.* 2015). In particular, when exploring patterns of genetic differentiation using only those loci considered to be outliers, marked genetic differentiation was revealed. Significantly, trout from the various clean rivers, separated by hundreds of kilometres, showed little to no genetic structuring between them, whereas the Hayle and Red River each formed distinctive

groups. This is indicative of differential patterns of local adaptation acting in each metal-impacted population to the unique cocktails of metals present in each river. In experimental evolution experiments, replicate populations often diverge from one another in their relative fitness and performance in other environments, suggesting that each population may be approaching different local peaks in the adaptive landscape (Elena & Lenski 2003). For example, populations of the euryhaline fish *Cottus asper* demonstrated patterns of repeated differentiation between brackish and freshwater ecotypes and associated high pairwise F_{ST} values (0.096-0.115), despite significant rates of gene flow between habitats (Dennenmoser *et al.* 2017) and strong genetic differentiation is known to occur in cave-dwelling fish *Poecilia mexicana* both between habitats and within pools in relation to toxic hydrogen sulphide (Plath *et al.* 2006).

Intriguingly, when using covarying markers of small effect associated with metal contamination, the Gannel population also formed a distinct group. Firstly, this suggests that the two main approaches used to find outlier loci differ in their detection specificity, as has been shown elsewhere (Benestan *et al.* 2016; Christmas *et al.* 2016). Whilst we cannot ascertain with confidence the reason for the Gannel population's distinctive genetic profile using LFMM outlier loci, we suggest two hypotheses. Firstly, the Gannel population, although initially selected as representing trout from a 'clean' river, could – given it's geographical location - be affected by contemporary, or at least, very recent metal pollution, potentially driving selection in this population. For instance, the Gannel has the second highest levels of iron of the five rivers studies here (Table 1). Secondly, the underlying metal-rich geology of the Cornwall region may have selected for low-frequency alleles involved in initial suitability of trout to the geological features of the metalliferous landscape. Selective forces driving metal-tolerance in brown trout populations experiencing contemporary elevated concentrations of metal pollution (including the Hayle and the Red River) may have acted on a large proportion of standing genetic variation (Barrett & Schluter 2008; Matuszewski *et al.* 2015). Moreover, the Gannel population showed the highest effective population size (Ne ~645) and could therefore potentially harbour a larger pool of ancestral low frequency 'metal-

type' genotypes, which have been subsequently acted upon in contemporary metal-tolerant populations.

Isolation-by-environment using the environmental distances between all metal contaminants present in each river was not significant. This can be explained by two key features of metal pollution. Firstly, the varying levels of metal pollution between each of the rivers could be masking the relative contribution of other metals, so, for example, high concentrations of arsenic in the Red River, likely concealed any patterns associated with more subtly varying concentrations of nickel. Secondly, this pattern could be attributed to the complexity of selection pressures imposed by metal pollution. As metals are classified as both essential for life (Cu, Zn, Ni) and non-essential (As, Cd), the biological handling of metals is complex. Thus, given the varying selection pressures each specific metal might impose, a straightforward pattern associating genetic distances with all metal pollutants would not be expected.

Arsenic represents one such non-essential metal, for which the LFMM analysis identified four unique covarying outlier loci. Chronic toxicity to arsenic is relatively slow acting, but it is known to affect growth, reproduction, immune response and osmoregulatory homeostasis (McIntyre & Linton 2012). Long-term toxicity of cadmium, on the other hand, is known to affect a wide array of processes including embryonic development, ionoregulation, energy metabolism, immunity and reproduction (McGeer *et al.* 2012). Toxic concentrations of essential metals, such as copper, are known to induce olfactory impairment and negatively affect ammonia excretion and acid-based balance (Grosell 2012) and chronic exposures involving zinc are known to primarily affect growth and reproduction (Hogstrand 2012). Furthermore, the metal concentration differences between the Hayle and the Red River varied in respect to specific pollutants (e.g. Hayle As ~15 µg/L vs. Red River As ~70 µg/L; Hayle Fe ~380 µg/L vs. Red River Fe ~3400 µg/L). The varying concentrations of these metals, as well as their different health and toxicity impacts, provide evidence for the patterns of IBE and genetic differentiation observed. Finally, this complexity of the biological handling of metals, and the varying concentrations of metals within each metal-impacted river, also points to a higher likelihood of unique and varying selection pressures within each metal-tolerant population of trout.

The relative contributions of parallel and convergent local adaptation between metal-impacted populations

Loci under divergent selection and regions tightly linked to them are expected to demonstrate strong patterns of differentiation across the genome, and associated with this, increased regions of homozygosity – 'islands of genomic divergence'. In the LHC plot, an illustration of such patterns were assessed in order to investigate the relative contributions of parallel and convergent local adaptation driving differences between the metal-impacted populations. As was also shown in the analyses above, the Hayle population showed higher levels of divergence compared to the Red River trout, suggesting stronger selection pressures acting on this population. Although both metal-impacted populations are known to have experienced metal pollution over a comparable amount of time, the recent mining activity of the 19th Century was particularly intense along the River Hayle, with dozens of mines existing across the catchment, extracting a wide variety of metals. This was particularly acute in the middle region of the river, 'Godolphin', where the Godolphin Mines operated adjacent to the river, extracting up to 8,425 tons of copper ore over approximately seven years (Dines 1956). Therefore, unique cocktails of metal pollutants have likely driven varying intensities of genomic divergence, suggesting differential strengths of selection operating on each metal-impacted population. For example, loci under selection for red and orange colour pattern elements in *Heliconius* butterflies were shown to illustrate stronger and more consistent patterns over larger regions of the genome, in comparison to loci under selection for yellow and white colour patterns (Nadeau *et al.* 2012). Furthermore, although patterns of divergence were occasionally reproduced in both populations, albeit significantly more in Hayle trout, a few specific chromosomal regions were differentially diverged across the two metal-impacted populations. In a review of heterogeneous genomic divergence, the majority of studies reported replicated divergence across population pairs (Nosil *et al.* 2009). However, parallel phenotypic divergence between ecotypes of *Littorina saxatilis* showed a low level of shared genomic divergence (~2–9%; Ravient *et al.* 2016) despite evidence of high gene flow (Butlin *et al.* 2013), and among four populations of *Arabidopsis halleri* growing in metal polluted or control habitats, the two metal-impacted populations showed different candidate markers, pointing to the

different roles of these loci in conferring metal tolerance (Meyer *et al.* 2009). Overall, these patterns suggest that both shared adaptive features, and regions of selection unique to each metal-impacted population likely contribute to the local adaptation of these populations to unique cocktails of metal contaminants.

Population-specific loci associated with gene ontology associated with traits that potentially confer ability to inhabit metal contaminated waters.

Contrasting mechanisms of metal-tolerance were again evidenced by the different sets of gene annotations that were observed for the various outlier loci. Interpretation of gene ontology should be assessed with caution, as stating ecological significance of ontology derived from investigations related to model species may be misguided. However, such functional annotation provides a valuable framework for the biological mechanisms responsible for selection (Bourret *et al.* 2013; Pavey *et al.* 2015), and potential areas for further research. In the F_{ST} outlier analyses comparing all populations equally, twelve-outlier loci strongly differentiated the Hayle and Red River samples from the three clean populations. Annotation of these outlier loci included homology to clathrin heavy chain 1 (CLTC) and a disintegrin and metalloproteinase domain-containing protein (ADAM8). ADAM8 has been induced in cadmium exposure of human blood mononuclear cells (Shin *et al.* 2003; Bakshi *et al.* 2008), CLTC was strongly upregulated in a Cd exposure of zebrafish (Wu *et al.* 2013) and clathrin-mediated endocytosis may be involved in mediating Cd^{2+} complexed to metallothionein (CdMT) (Carvou *et al.* 2007*;* Thévenod 2010).

A locus showing homology to antizyme inhibitor 1 was consistently identified as an outlier in both the Hayle and the Red River. Antizyme inhibitors positively regulate ornithine decarboxylase (ODC) activity and polyamine uptake. ODC is an enzyme that catalyses the biosynthesis of polyamines (Pegg 2006), which have a wide range of functions in cell growth and division (Heby & Persson 1990), but of interest here, are also implicated as having key functions in ionoregulation and regulation of osmolality (Poulin & Pegg 1990). Polyamine levels have been shown to be important in stabilising genome function in the face of rapid changes in osmolyte concentrations during osmotic challenge in fish (Whitehead *et al.* 2011; Whitehead *et al.* 2012; Leguen *et al.* 2015; Guan *et al.* 2016) and ODC regulation has been associated with cellular apoptosis and mortality in sockeye salmon (*Oncorhynchus nerka*), likely associated with

osmoregulatory failure (Jeffries *et al.* 2012). Moreover, it has recently been shown that increased expression of AZIN1 is related to dehydration, salt loading and acute hypertonic stress and *in vivo* analysis silencing the expression of AZIN1 decreased plasma osmolality, and also had a significant effect on food intake (Greenwood *et al.* 2015). Given the large impact that metals have on iono- and osmo- regulation in fish (Paris *et al.* Chapter 5), this gene could play a key role in adjusting the homeostasis of osmoregulation in metal-impacted trout populations.

Outlier loci identified uniquely in the Hayle population included a locus with homology to ras-like protein family member 10B. Ras proteins are a group of GTPase switches that function in intracellular signalling pathways. Mutant forms of Ras proteins have been associated with several types of mammalian cancer (Bos 1989; Schubbert *et al.* 2007), yet in fish, patterns of ras gene regulation and modification in response to environmental contaminants appears to be complex (Rotchell *et al.* 2001). Another outlier was identified as similar to ephrin type-A receptor 7, which may be involved in altering gill architecture in response to osmotic stress (Pinto *et al.* 2010; Lam *et al.* 2014; Lai *et al.* 2015). Furthermore, analysis of the GO terms associated with the Hayle F_{ST} outliers showed significant enrichment of gene annotations involved in the positive regulation of GTPase activity. Amongst their many biological functions, GTPases play a role in cytoskeleton arrangement (Sit & Manser 2011) and may represent major adaptive responses and genetic reprogramming of cytoskeleton remodelling under hypo-osmotic conditions (Di Ciano-Oliveira *et al.* 2006), and may also play a role in responses to metals (Devos *et al.* 1998; Eyckmans *et al.* 2012). It is therefore possible that selection may be acting on GTPases in order to alter gill cytoskeleton structure.

Analysis of gene ontology of the F_{ST} outliers under selection specifically in trout from the Red River included ontology to lamina-associated polypeptide 2 (LAP2) and SUN domain-containing ossification factor-like (SUCO). LAP2 (also known as thymopoietin) has been implicated in mammalian T-cell development (Kosak *et al.* 2002; Krangel 2007) and was upregulated in the gills of vaccinated salmon challenged by bacterial infection (Bridle *et al.* 2012). LAP2 may also be involved in chromosome organisation (Harris *et al.* 1995; Zheng *et al.* 2000); and a variety of metals are known to cause DNA breakage and chromosomal

abnormalities (Moore 1985; da Silva Souza *et al.* 2006; Ergene *et al.* 2007). SUCO is a developmental protein required for bone modelling, and was upregulated in the gill of Japanese flounder (*Paralichthys olivaceus*) exposed to crude oil (Zhu *et al.* 2016) and in phosphorus deficient rainbow trout (*O. mykiss*) (Le Luyer *et al.* 2015). GO analysis of the Red River outliers showed significant enrichment of receptor signalling protein tyrosine kinase activity. Receptor tyronisine kinases (RTKs) regulate several aspects of cell proliferation and differentiation, cell survival, and cellular metabolism. A variety of stressors are known to stimulate tyrosine kinase activity. Heavy metals have been shown to produce specific tyrosine kinase responses in trout cell exposure (Burlando *et al.* 2003) and may be important in arsenic-induced responses of fish (Bhattacharya *et al.* 2007).

Lastly, the LFMM analysis showed that one RAD locus, which showed homology to RNA binding protein 4 (RBM4), was under selection associated with arsenic, nickel and zinc. RBM4 is involved in multiple aspects of cellular processes, including alternative splicing of pre-mRNA and translation regulation (Lin & Tarn 2011; Catsello *et al.* 2012). RBM4 is known to exert suppressive activity on biological processes that are increased under stress conditions (Alves & Goldenberg 2016), including arsenic (Lin *et al.* 2007). Together, as key regulators of gene expression, and the role these proteins play in regulating stress responses, the identification of this gene as an outlier related specifically to certain metals may represent an important modulator of adaptation to metal pollution in these trout populations.

De novo and standing genetic variation in driving metal-tolerant trout populations

A major question in adaptation to novel environments is whether selection acts on standing genetic variation or as the result of *de novo* mutations (Barrett & Schluter 2008; Elmer & Meyer 2001; Messer & Petrov 2013). For example, rapid colonisation of freshwater niches in the threespine stickleback (*Gasterosteus aculeatus*) is facilitated by fixation of loci already present in the oceanic ecotypes (Hohenlohe *et al* 2010; Feulner *et al.* 2013). Detecting these different modes of selection may result in different patterns of hard sweeps and soft sweeps. Here, we observed preliminary evidence for both processes. Fixation of diversity in a handful of F_{ST} outlier loci in metal-impacted populations

suggests selection to metals from standing genetic variation, yet, increased diversity at other loci suggests the role of *de novo* mutations in conferring metal-tolerant traits. The patterns of differentiation and haplotype diversity in the LHC plot point to polygenic standing genetic variation. In particular, rapid adaptation to novel environments is expected to result from standing genetic variation, and signals of such selection are likely to be restricted to shorter chromosomal regions (Barret & Schluter 2008; Counterman *et al.* 2010). Exploration of differentiation in the American eel (*Anguilla rostrata*) revealed that 331 SNPs distributed across 325 different scaffolds distinguished between freshwater and brackish/saltwater 'modules', (Pavey *et al.* 2015) and 141 polygenic SNPs differentiated the species between control and polluted sampling localities (Laporte *et al.* 2016). Lastly, the detection of different subsets of loci under the contrasting selection tests suggests that different regions of the genome may be involved. Future research should aim to assess whether such polymorphisms result in synonymous or non-synonymous mutations.

Concluding remarks

Quantifying the response of local populations is paramount in assessing the role of genetic diversity in conferring local adaptation to human-altered environments. The main aim of this study was to assess the role of metal pollutants in driving local adaptation of brown trout populations to unique cocktails of metals. Our results indicate that trout have diverged genetically in each metal-impacted habitat despite the close spatial proximity of the two rivers, which is consistent with differential drivers of selection at a local scale and correspondingly different patterns of local adaptation. Although we should consider that the strong genetic signals are indicative of a combination of both neutral and non-neutral genetic forces, the identification of a suite of both F_{ST} and covarying outlier loci related specifically to each population, and correlated with particular metal contaminants confirms that local adaptation is an important process acting in these populations. Finally, annotation of outlier loci and identification of islands of divergence between the populations also lends significant support to the hypothesis that these populations are locally adapted. The genetic divergence and indicators of different mechanisms of local

adaptation suggests that these metal-impacted populations should be considered unique metal ecotypes within the brown trout species.

ACKNOWLEDGEMENTS

We would like to thank the University of Exeter, Environment Agency and Westcountry Rivers Trust for funding this work. Thanks to Matt Healey, Giles Ricard and Dr Bruce Stockley from the Westcountry Rivers Trust and also to Jan and Phil Shears for organising the electrofishing and assisting with fieldwork for the fish collection. Thank you to the University of Exeter Sequencing Service for assistance with preparation and sequencing of the RADseq libraries. Thanks also to Dr Pat Hamilton and Tom Jenkins for assistance and ideas for data analysis.

REFERENCES

Allendorf, F. W., & Hard, J. J. (2009). Human-induced evolution caused by unnatural selection through harvest of wild animals. *Proceedings of the National Academy of Sciences of the United States of America*, (Supplement 1), 9987–94.

Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews Genetics*, **11**(10), 697–709.

Alves, L. R., & Goldenberg, S. (2016). RNA-binding proteins related to stress response and differentiation in protozoa. *World Journal of Biological Chemistry*, **7**(1), 78-87.

Andolfatto, P. (2001). Adaptive hitchhiking effects on genome variability. *Current Opinion in Genetics & Development*, **11**(6), 635–641.

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, **17**(2), 81–92.

Antao, T., Lopes, A., Lopes, R. J., Beja-Pereira, A., & Luikart, G. (2008). LOSITAN: A workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics*, **9**(323), **DOI:** 10.1186/1471-2105-9-323

Antonovics, J., Bradshaw, A. D., & Turner, R. G. (1971). Heavy metal tolerance in plants. *Advances in Ecological Research*, **7**, 1–85.

Bakshi, S., Zhang, X., Godoy-Tundidor, S., Cheng, R. Y. S., Sartor, M. A., Medvedovic, M., & Ho, S.-M. (2008). Transcriptome analyses in normal

prostate epithelial cells exposed to low-dose cadmium: oncogenic and immunomodulations involving the action of tumor necrosis factor. *Environmental Health Perspectives*, **116**(6), 769–776.

Barrett, R. D. H., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, **23**(1), 38–44.

Barton, N. H. (2000). Genetic hitchhiking. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **355**(1403), 1553–62.

Beaumont, M. A., & Nichols, R. A. (1996). Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London B: Biological Sciences*, **263**(1377), 1619-1626.

Benestan, L., Quinn, B. K., Maaroufi, H., Laporte, M., Clark, F. K., Greenwood, S. J., Rochette, R., & Bernatchez, L. (2016). Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (*Homarus americanus*). *Molecular Ecology*, **25**(20), 5073–5092.

Benjamini Y & Hochberg Y (1995) Controlling the false discovery rate – a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*: *Series B (Statistical Methodology)* **57**, 289–300.

Bernatchez, L. (2001). The evolutionary history of brown trout (*Salmo trutta* L.) inferred from phylogeographic, nested clade, and mismatch analyses of mitochondrial DNA variation. *Evolution*, **55**(2), 351–379.

Bernatchez, L. (2016). On the maintenance of genetic variation and adaptation to environmental change: considerations from population genomics in fishes. *Journal of Fish Biology*, **89**: 2519–2556.

Bhattacharya, S., Bhattacharya, A., & Roy, S. (2007). Arsenic-induced responses in freshwater teleosts. *Fish Physiology and Biochemistry*, **33**(4), 463–473.

Bickham, J. W., Sandhu, S., Hebert, P. D., Chikhi, L., & Athwal, R. (2000). Effects of chemical contaminants on genetic diversity in natural populations: implications for biomonitoring and ecotoxicology. *Mutation Research*, **463**(1), 33–51.

Blanquart, F., Kaltz, O., Nuismer, S. L., & Gandon, S. (2013). A practical guide to measuring local adaptation. *Ecology Letters*, **16**(9), 1195–1205.

Bos, J. L. (1989). Ras oncogenes in human cancer: a review. *Cancer Research*, **49**(17), 4682–9.

Bougas, B., Normandeau, E., Grasset, J., Defo, M. A., Campbell, P. G. C., Couture, P., & Bernatchez, L. (2016). Transcriptional response of yellow

perch to changes in ambient metal concentrations—A reciprocal field transplantation experiment. *Aquatic Toxicology*, **173**, 132–142.

Bourret, V., Couture, P., Campbell, P. G. C., & Bernatchez, L. (2008). Evolutionary ecotoxicology of wild yellow perch (*Perca flavescens*) populations chronically exposed to a polymetallic gradient. *Aquatic Toxicology*, **86**(1), 76–90.

Bourret, V., Dionne, M., Kent, M. P., Lien, S., & Bernatchez, L. (2013). Landscape genomics in Atlantic salmon (*Salmo salar*): searching for gene-environment interactions driving local adaptation. *Evolution*, **67**(12), 3469–3487.

Bridle, A. R., Koop, B. F., Nowak, B. F., Walker, N., & Lieph, R. (2012). Identification of surrogates of protection against Yersiniosis in immersion vaccinated Atlantic salmon. *PLoS ONE*, **7**(7), e40841.

Britton-Davidian, J., Catalan, J., da Graça Ramalhinho, M., Ganem, G., Auffray, J.-C., Capela, R., Biscoito, M., Searle, J.B., & da Luz Mathias, M. (2000). Rapid chromosomal evolution in island mice. *Nature*, **403**(6766), 158–158.

Brown, B. (1977). Effects of mine drainage on the River Hayle, Cornwall a) factors affecting concentrations of copper, zinc and iron in water, sediments and dominant invertebrate fauna. *Hydrobiologia*, **52**, 2–3.

Bryan, G.W. and Gibbs, P.E. (1983) *Heavy metals in the Fal estuary, Cornwall: A study of long term contamination by mining waste and its effects on estuarine organisms.* Occasional Publications. Marine Biological Association of the United Kingdom.

Burlando, B., Magnelli, V., Panfoli, I., Berti, E., & Viarengo, A. (2003). Ligand-independent tyrosine kinase signalling in RTH 149 trout hepatoma cells: comparison among heavy metals and pro-oxidants. *Cellular Physiology and Biochemistry : International Journal of Experimental Cellular Physiology, Biochemistry, and Pharmacology*, **13**(3), 147–54.

Butlin, R. K., Saura, M., Charrier, G., Jackson, B., André, C., Caballero, A., Coyne J.A. *et al.* (2014). Parallel evolution of local adaptation and reproductive isolation in the face of gene flow. *Evolution*, **68**(4), 935–949.

Campbell, D., & Bernatchez, L. (2004). Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Molecular Biology and Evolution*, **21**(5), 945–956.

Carlsson, J., & Nilsson, J. (2000). Population genetic structure of brown trout (*Salmo trutta* L.) within a northern boreal forest stream. *Hereditas*, **132**(3), 173–81.

Carlsson, J., & Nilsson, J. (2001). Effects of geomorphological structures on genetic differentiation among brown trout populations in a northern boreal river drainage. *Transactions of the American Fisheries Society*, **130**(1), 36–45.

Carlsson, J., Olsen, K. H., Nilsson, J., Overli, O., & Stabell, O. B. (1999). Microsatellites reveal fine-scale genetic structure in stream-living brown trout. *Journal of Fish Biology*, **55**(6), 1290–1303.

Carvou, N., Norden, A. G. W., Unwin, R. J., & Cockcroft, S. (2007). Signalling through phospholipase C interferes with clathrin-mediated endocytosis. *Cellular Signalling*, **19**(1), 42–51.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B. M., Strein, C., *et al.* (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell*, **149**(6), 1393–1406.

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks: building and genotyping Loci de novo from short-read sequences. *G3,* **1**(3), 171–82.

Catchen, J., Hohenlohe, P. a., Bassham, S., Amores, A., & Cresko, W.A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, **22**(11), 3124-40.

Charlesworth, B. (1996). Background selection and patterns of genetic diversity in *Drosophila melanogaster*. *Genetical Research*, **68**(2), 131–49.

Charlesworth, B. (2012). The role of background selection in shaping patterns of molecular evolution and variation: evidence from variability on the Drosophila X chromosome. *Genetics*, **191**(1), 233–246.

Chhatre, V. E., & Emerson, K. J. (2017). StrAuto: automation and parallelization of STRUCTURE analysis. *BMC Bioinformatics*, **18**(192), **DOI:** 10.1186/s12859-017-1593-0

Christmas, M. J., Biffin, E., Breed, M. F., & Lowe, A. J. (2016). Finding needles in a genomic haystack: targeted capture identifies clear signatures of selection in a nonmodel plant species. *Molecular Ecology*, **25**(17), 4216–4233.

Counterman, B. A., Araujo-Perez, F., Hines, H. M., Baxter, S. W., Morrison, C. M., Lindstrom, D. P. *et al.* (2010). Genomic hotspots for adaptation: the population genetics of Müllerian mimicry in *Heliconius erato*. *PLoS Genetics*, **6**(2), e1000796.

Colson, I., & Hughes, R. N. (2004). Rapid recovery of genetic diversity of dogwhelk (*Nucella lapillus* L.) populations after local extinction and recolonization contradicts predictions from life-history characteristics. *Molecular Ecology*, **13**(8), 2223–2233.

Crutzen, P. J. (2002). Geology of mankind. *Nature*, **415**(6867), 23–23.

Cutter, A. D., & Payseur, B. A. (2013). Genomic signatures of selection at linked sites: unifying the disparity among species. *Nature Reviews Genetics*, **14**(4), 262–274.

da Silva Souza, T., & Fontanetti, C. S. (2006). Micronucleus test and observation of nuclear alterations in erythrocytes of Nile tilapia exposed to waters affected by refinery effluent. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, **605**(1-2), 87–93.

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, **12**(7), 499–510.

De La Vega-Salazar, M. Y., Avila-Luna, E., & Macías-Garcia, C. (2003). Ecological evaluation of local extinction: the case of two genera of endemic Mexican fish, Zoogoneticus and Skiffia. *Biodiversity and Conservation*, **12**(10), 2043–2056.

De Mita, S., Thuillet, A.-C., Gay, L., Ahmadi, N., Manel, S., Ronfort, J., & Vigouroux, Y. (2013). Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular Ecology*, **22**(5), 1383–1399.

Dennenmoser, S., Vamosi, S. M., Nolte, A. W., & Rogers, S. M. (2017). Adaptive genomic divergence under high gene flow between freshwater and brackish-water ecotypes of prickly sculpin (*Cottus asper*) revealed by Pool-Seq. *Molecular Ecology*, **26**(1), 25–42.

Devos, E., Devos, P., & Cornet, M. (1998). Effect of cadmium on the cytoskeleton and morphology of gill chloride cells in parr and smolt Atlantic salmon (*Salmo salar*). *Fish Physiology and Biochemistry*, **18**(1), 15–27.

Di Ciano-Oliveira, C., Thirone, A. C. P., Szaszi, K., & Kapus, A. (2006). Osmotic stress and the cytoskeleton: the R(h)ole of Rho GTPases. *Acta Physiologica*, **187**(1-2), 257–272.

Dines, H. G. (1956). *The Metalliferous Mining Region of South-West England, Vol. 2.* H.M. Stationery Office, *London*.

Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J., & Ovenden, J. R. (2014). NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (N_e) from genetic data. *Molecular Ecology Resources*, **14**(1), 209–214.

Duguid, R. A., Ferguson, A., & Prodohl, P. (2006). Reproductive isolation and genetic differentiation of ferox trout from sympatric brown trout in Loch Awe and Loch Laggan, Scotland. *Journal of Fish Biology*, **69**, 89–114.

Durrant, C. J., Stevens, J. R., Hogstrand, C., & Bury, N. R. (2011). The effect of metal pollution on the population genetic structure of brown trout (*Salmo trutta* L.) residing in the River Hayle, Cornwall, UK. *Environmental Pollution,* **159**(12), 3595–603.

Dutoit, L., Burri, R., Nater, A., Mugal, C. F., & Ellegren, H. (2016). Genomic distribution and estimation of nucleotide diversity in natural populations: perspectives from the collared flycatcher (*Ficedula albicollis*) genome. *Molecular Ecology Resources*. DOI: 10.1111/1755-0998.12602

Elena, S. F., & Lenski, R. E. (2003). Microbial genetics: Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nature Reviews Genetics*, **4**(6), 457–469.

Ellis, J. S., Sumner, K. J., Griffiths, A. M., Bright, D. I., & Stevens, J. R. (2011). Population genetic structure of Atlantic salmon, *Salmo salar* L., in the River Tamar, southwest England. *Fisheries Management and Ecology*, **18**(3), 233–245.

Elmer, K. R., & Meyer, A. (2011). Adaptation in the age of ecological genomics: insights from parallelism and convergence. *Trends in Ecology & Evolution*, **26**(6), 298–306.

Environment Agency (2013). Abandoned mines and the water environment. Science Project: SC030136-41.

Ergene, S., Çavaş, T., Çelik, A., Köleli, N., Kaya, F., & Karahan, A. (2007). Monitoring of nuclear abnormalities in peripheral erythrocytes of three fish species from the Goksu Delta (Turkey): genotoxic damage in relation to water pollution. *Ecotoxicology*, **16**(4), 385–391.

Etter, P. D., Preston, J. L., Bassham, S., Cresko, W.A., & Johnson, E.A. (2011). Local de novo assembly of RAD paired-end contigs using short sequencing reads. *PloS One*, **6**(4), e18561.

Evanno, G., Regnaut, S. and Goudet, J. (2005), Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier, L., Hofer, T., & Foll, M. (2009). Detecting loci under selection in a hierarchically structured population. *Heredity*, **103**(4), 285–298.

Eyckmans, M., Benoot, D., Van Raemdonck, G. A. A., Zegels, G., Van Ostade, X. W. M., Witters, E. *et al.* (2012). Comparative proteomics of copper exposure and toxicity in rainbow trout, common carp and gibel carp. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, **7**(2), 220–232.

Ferguson, A., & Mason, F. M. (1981). Allozyme evidence for reproductively isolated sympatric populations of brown trout *Salmo trutta* L. in Lough Melvin, Ireland. *Journal of Fish Biology*, **18**(6), 629–642.

Feulner, P. G. D., Chain, F. J. J., Panchal, M., Eizaguirre, C., Kalbe, M., Lenz, T. L. *et al.* (2013). Genome-wide patterns of standing genetic variation in a marine population of three-spined sticklebacks. *Molecular Ecology*, **22**(3), 635–649.

Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**(2), 977-93.

Francis, R. M. (2017), pophelper: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, **17**, 27–32.

Fraser, D. J., Hansen, M. M., Østergaard, S., Tessier, N., Legault, M., & Bernatchez, L. (2007). Comparative estimation of effective population sizes and temporal gene flow in two contrasting population systems. *Molecular Ecology*, **16**(18), 3866–3889.

Fraser, D. J., Weir, L. K., Bernatchez, L., Hansen, M. M., & Taylor, E. B. (2011). Extent and scale of local adaptation in salmonid fishes: review and meta-analysis. *Heredity*, **106**(3), 404–420.

Frichot, E., & François, O. (2015). LEA : An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, **6**(8), 925–929.

Frichot, E., Schoville, S. D., Bouchard, G., & François, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, **30**(7), 1687–99.

Gillespie, J. H. (1997). Junk ain't what junk does: neutral alleles in a selected context. *Gene*, **205**(1), 291–299.

Goetz, F., Rosauer, D., Sitar, S., Goetz, G., Simchick, C., Roberts, S., *et al.* (2010). A genetic basis for the phenotypic differentiation between siscowet and lean lake trout (*Salvelinus namaycush*). *Molecular Ecology*, **19**, 176–196.

Gonzalez-Quevedo, C., Spurgin, L. G., Illera, J. C., & Richardson, D. S. (2015). Drift, not selection, shapes toll-like receptor variation among oceanic island populations. *Molecular Ecology*, **24**(23), 5852–5863.

Gordo, I., Navarro, A., & Charlesworth, B. (2002). Muller's ratchet and the pattern of variation at a neutral locus. *Genetics*, **161**(2), 835-48.

Goslee, S. C., & Urban, D. L. (2007). The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, **22**(7), 1–19.

Gray, D., Harding, J., & Lindsay, P. (2016). Remediation of a major acid mine drainage point source discharge restores headwater connectivity for a diadromous native fish. *New Zealand Journal of Marine and Freshwater Research*, **50**(4), 566–580.

Greenwood, M. P., Greenwood, M., Paton, J. F. R., & Murphy, D. (2015). Control of polyamine biosynthesis by antizyme inhibitor 1 is important for transcriptional regulation of arginine vasopressin in the male rat hypothalamus. *Endocrinology*, **156**(8), 2905–2917.

Griffiths, A. M., Koizumi, I., Bright, D., & Stevens, J. R. (2009). A case of isolation by distance and short-term temporal stability of population structure in brown trout (*Salmo trutta*) within the River Dart, southwest England. *Evolutionary Applications*, **2**(4), 537–554.

Griffiths, A. M., Machado-Schiaffino, G., Dillane, E., Coughlan, J., Horreo, J. L., Bowkett, A. E., *et al.* (2010). Genetic stock identification of Atlantic salmon (*Salmo salar*) populations in the southern part of the European range. *BMC Genetics*, **11**(31). DOI: 10.1186/1471-2156-11-31.

Grivet, D., Sebastiani, F., Alia, R., Bataillon, T., Torre, S., Zabal-Aguirre, M. *et al.* (2011). Molecular footprints of local adaptation in two Mediterranean conifers. *Molecular Biology and Evolution*, **28**(1), 101–116.

Grosell, M. (2012). Copper. In: C.M. Wood, A.P. Farrell & C.J. Brauner (eds.), *Homeostasis and toxicology of essential metals*, Volume 31A. Cambridge, MA: Elsevier Academic Press, pp. 54-110.

Guan, Y., Zhang, G., Zhang, S., Domangue, B., & Galvez, F. (2016). The potential role of polyamines in gill epithelial remodeling during extreme hypo-osmotic challenges in the Gulf killifish, *Fundulus grandis*. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, **194**, 39–50.

Haasl, R. J., & Payseur, B. A. (2016). Fifteen years of genomewide scans for selection: trends, lessons and unaddressed genetic sources of complication. *Molecular Ecology*, **25**(1), 5–23.

Hadfield, J. D. (2016). The spatial scale of local adaptation in a stochastic environment. *Ecology Letters*, **19**(7), 780–788.

Hairston, N. G., Ellner, S. P., Geber, M. A., Yoshida, T., & Fox, J. A. (2005). Rapid evolution and the convergence of ecological and evolutionary time. *Ecology Letters*, **8**(10), 1114–1127.

Hamilton, P. B., Cowx, I. G., Oleksiak, M. F., Griffiths, A. M., Grahn, M., Stevens, J. R. *et al.* (2016). Population-level consequences for wild fish exposed to sublethal concentrations of chemicals - a critical review. *Fish and Fisheries*, **17**(3), 545–566.

Hansen, M. M., Ruzzante, D. E., Nielsen, E. E., Bekkevold, D., & Mensberg, K.-L. D. (2002). Long-term effective population sizes, temporal stability of genetic composition and potential for local adaptation in anadromous brown trout (*Salmo trutta*) populations. *Molecular Ecology*, **11**(12), 2523–2535.

Harris, C. A., Andryuk, P. J., Cline, S. W., Mathew, S., Siekierka, J. J., & Goldstein, G. (1995). Structure and mapping of the human thymopoietin (TMPO) gene and relationship of human TMPO β to rat lamin-associated polypeptide 2. *Genomics*, **28**(2), 198–205.

Heby, O., & Persson, L. (1990). Molecular genetics of polyamine synthesis in eukaryotic cells. *Trends in Biochemical Sciences*, **15**(4), 153–158.

Hendry, A. P., & Kinnison, M. T. (2001). An introduction to microevolution: rate, pattern, process. *Genetica*, **112/113**(1), 1–8.

Hereford, J. (2009). A quantitative survey of local adaptation and fitness trade-offs. *The American Naturalist*, **173**(5), 579–88.

Hermisson, J., & Pennings, P. S. (2005). Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics*, **169**(4), 2335-52.

Hermisson, J., & Pennings, P. S. (2017). Soft sweeps and beyond: Understanding the patterns and probabilities of selection footprints under rapid adaptation. *bioRxiv*, DOI: https://doi.org/10.1101/114587

Hess, J. E., Campbell, N. R., Close, D. A., Docker, M. F., & Narum, S. R. (2013). Population genomics of Pacific lamprey: adaptive variation in a highly dispersive species. *Molecular Ecology*, **22**(11), 2898–2916.

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B. *et al.* (2016). Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *The American Naturalist*, **188**(4), 379–97.

Hogstrand, C. (2012). Zinc. In: C.M. Wood, A.P. Farrell & C.J. Brauner (eds.), *Homeostasis and toxicology of essential metals*, Volume 31A. Cambridge, MA: Elsevier Academic Press, pp. 136-184.

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010). Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. *PLoS Genetics*, **6**(2), e1000862.

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, **4**(1), 44–57.

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*, **37**(1), 1–13.

Jeffries, K. M., Hinch, S. G., Sierocinski, T., Clark, T. D., Eliason, E. J., Donaldson, M. R. *et al.* (2012). Consequences of high temperatures and premature mortality on the transcriptome and blood physiology of wild adult sockeye salmon (*Oncorhynchus nerka*). *Ecology and Evolution*, **2**(7), 1747–1764.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**(11), 1403–1405.

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27**(21), 3070–3071.

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, **11**(94), DOI: 10.1186/1471-2156-11-94.

Kamvar, Z. N., Brooks, J. C., & Grünwald, N. J. (2015). Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Frontiers in Genetics*, **6**, 208.

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). *Poppr*: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, **2**, e281.

Kanno, Y., Vokoun, J. C., & Letcher, B. H. (2011). Fine-scale population structure and riverscape genetics of brook trout (*Salvelinus fontinalis*) distributed continuously along headwater channel networks. *Molecular Ecology*, **20**(18), 3711–3729.

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, **7**(12), 1225–1241.

Keeley, E. R., Parkinson, E. A., & Taylor, E. B. (2007). The origins of ecotypic variation of rainbow trout: a test of environmental vs. genetically based differences in morphology. *Journal of Evolutionary Biology*, **20**(2), 725–736.

Kennedy, E. V., Tonk, L., Foster, N. L., Chollett, I., Ortiz, J.-C., Dove, S *et al.* (2016). *Symbiodinium* biogeography tracks environmental patterns rather than host genetics in a key Caribbean reef-builder, *Orbicella annularis*. *Proceedings of the Royal Society B: Biological Sciences*, **283**(1842), 20161938.

King, R. A., Hillman, R., Elsmere, P., Stockley, B., & Stevens, J. R. (2016). Investigating patterns of straying and mixed stock exploitation of sea trout,

Salmo trutta, in rivers sharing an estuary in south-west England. *Fisheries Management and Ecology*, **23**(5), 376–389.

Kinnison, M. T., & Hendry, A. P. (2001). The pace of modern life II: From rates of contemporary microevolution to pattern and process. *Genetica*, **112-113**, 145-64.

Kosak, S. T., Skok, J. A., Medina, K. L., Riblet, R., Le Beau, M. M., Fisher, A. G., & Singh, H. (2002). Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science*, **296**(5565), 158-162.

Krangel, M. S. (2007). T cell development: better living through chromatin. *Nature Immunology*, **8**(7), 687–694.

Lai, K. P., Li, J.-W., Gu, J., Chan, T.-F., Tse, W. K. F., & Wong, C. K. C. (2015). Transcriptomic analysis reveals specific osmoregulatory adaptive responses in gill mitochondria-rich cells and pavement cells of the Japanese eel. *BMC Genomics*, **16**, 1072. DOI: 10.1186/s12864-015-2271-0.

Lam, S., Lui, E., Li, Z., Cai, S., Sung, W.-K., Mathavan, S., *et al.* (2014). Differential transcriptomic analyses revealed genes and signaling pathways involved in iono-osmoregulation and cellular remodeling in the gills of euryhaline Mozambique tilapia, *Oreochromis mossambicus*. *BMC Genomics*, **15**(921), DOI: 10.1186/1471-2164-15-921.

Lande, R. (1998). Anthropogenic, ecological and genetic factors in extinction and conservation. *Researches on Population Ecology*, **40**(3), 259–269.

Lande, R., & Shannon, S. (1996). The role of genetic variation in adaptation and population persistence in a changing environment. *Evolution*, **50**(1), 434-437.

Landry, L., Vincent, W. F., & Bernatchez, L. (2007). Parallel evolution of lake whitefish dwarf ecotypes in association with limnological features of their adaptive landscape. *Journal of Evolutionary Biology*, **20**(3), 971–984.

Laporte, M., Pavey, S. A., Rougeux, C., Pierron, F., Lauzent, M., Budzinski, H. *et al.* (2016). RAD sequencing reveals within-generation polygenic selection in response to anthropogenic organic and metal contamination in North Atlantic Eels. *Molecular Ecology*, **25**(1), 219–237.

Larsen, P. F., Nielsen, E. E., Williams, T. D., Hemmer-Hansen, J., Chipman, J. K., Kruhøffer, M. *et al.* (2007). Adaptive differences in gene expression in European flounder (*Platichthys flesus*). *Molecular Ecology*, **16**(22), 4674–4683.

Lee, C. E., & MitchelL-Olds, T. (2006). Preface to the special issue: ecological and evolutionary genomics of populations in nature. *Molecular Ecology*, **15**(5), 1193–1196.

Leguen, I., Le Cam, A., Montfort, J., Peron, S., & Fautrel, A. (2015). Transcriptomic analysis of trout gill ionocytes in freshwater and seawater using laser capture microdissection combined with microarray analysis. *PLOS ONE*, **10**(10), e0139938.

Lescak, E. A., Bassham, S. L., Catchen, J., Gelmond, O., Sherbick, M. L., von Hippel, F. A., & Cresko, W.A. (2015). Evolution of stickleback in 50 years on earthquake-uplifted islands. *Proceedings of the National Academy of Sciences of the United States of America*, **112**(52), E7204–12.

Lin, J.-C., & Tarn, W.-Y. (2011). RBM4 down-regulates PTB and antagonizes its activity in muscle cell–specific alternative splicing. *The Journal of Cell Biology*, **193**(3), 509–520.

Lin, J.-C., Hsu, M., & Tarn, W.-Y. (2007). Cell stress modulates the function of splicing regulatory protein RBM4 in translation control. *Proceedings of the National Academy of Sciences*, **104**(7), 2235–2240.

Lischer, H. E. L., & Excoffier, L. (2012) PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics,* **28**: 298-299.

Lotterhos, K. E. and Whitlock, M. C. (2015), The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology*, **24**, 1031–1046.

Le Luyer, J., Deschamps, M.H., Proulx, E., Poirier Stewart, N., Droit, A., Sire, J.Y., Robert, C., & Vandenberg, G.W. (2015). RNA-Seq transcriptome analysis of pronounced biconcave vertebrae: A common abnormality in rainbow trout (*Oncorhynchus mykiss*) fed a low-phosphorus diet. *Journal of Next Generation Sequencing & Applications*, **2**, 112, DOI:10.4172/2469-9853.1000112.

Marques, D. A., Lucek, K., Meier, J. I., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2016). Genomics of rapid incipient speciation in sympatric threespine stickleback. *PLOS Genetics*, **12**(2), e1005887.

Matuszewski, S., Hermisson, J., & Kopp, M. (2015). Catch me if you can: adaptation from standing genetic variation to a moving phenotypic optimum. *Genetics*, **200**(4), 1255–74.

Maynard Smith, J., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical Research*, **23**(1), 23-25.

McGeer, J.C., Niyogi, S., & Smith, D.S. (2012). Cadmium. In: C.M. Wood, A.P. Farrell & C.J. Brauner (eds.), *Homeostasis and toxicology of non-essential metals*, Volume 31B. Cambridge, MA: Elsevier Academic Press, pp. 126-168.

McIntyre & Linton (2012). Arsenic. In: C.M. Wood, A.P. Farrell & C.J. Brauner (eds.), *Homeostasis and toxicology of non-essential metals*, Volume 31B. Cambridge, MA: Elsevier Academic Press, pp. 298-337.

McKeown, N. J., Hynes, R. A., Duguid, R. A., Ferguson, A., & Prodöhl, P. A. (2010). Phylogeographic structure of brown trout *Salmo trutta* in Britain and Ireland: glacial refugia, postglacial colonization and origins of sympatric populations. *Journal of Fish Biology*, **76**(2), 319–347.

Meier, K., Hansen, M. M., Bekkevold, D., Skaala, Ø., & Mensberg, K.-L. D. (2011). An assessment of the spatial scale of local adaptation in brown trout (*Salmo trutta* L.): footprints of selection at microsatellite DNA loci. *Heredity*, **106**(3), 488–499.

Meirmans, P. G., & van Tienderen, P. H. (2004). Genotype and Genodive: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, **4**(4), 792–794.

Messer, P. W., & Petrov, D. A. (2013). Population genomics of rapid adaptation by soft selective sweeps. *Trends in Ecology & Evolution*, **28**(11), 659–669.

Messer, P. W., Ellner, S. P., & Hairston, N. G. (2016). Can population genetics adapt to rapid evolution? *Trends in Genetics*, **32**(7), 408–418.

Meyer, C.-L., VItalis, R., Saumitou-Laprade, P., & Castric, V. (2009). Genomic pattern of adaptive divergence in *Arabidopsis halleri*, a model species for tolerance to heavy metal. *Molecular Ecology*, **18**(9), 2050–2062.

Miller, M. R., Brunelli, J. P., Wheeler, P.A., Liu, S., Rexroad, C. E., Palti, Y. *et al.* (2012). A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. *Molecular Ecology*, **21**(2), 237–49.

Minghetti, M., Schnell, S., Chadwick, M. A., Hogstrand, C., & Bury, N. R. (2014). A primary FIsh Gill Cell System (FIGCS) for environmental monitoring of river waters. *Aquatic Toxicology*, **154**, 184–192.

Moore, M. N. (1985). Cellular responses to pollutants. *Marine Pollution Bulletin*, **16**(4), 134–139.

Morjan, C. L., & Rieseberg, L. H. (2004). How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Molecular Ecology*, **13**(6), 1341–1356.

Nadeau, N. J., Whibley, A., Jones, R. T., Davey, J. W., Dasmahapatra, K. K., Baxter, S. W. *et al.* (2011). Genomic islands of divergence in hybridizing Heliconius butterflies identified by large-scale targeted sequencing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**(1587), 343-53.

Narum, S. R., & Hess, J. E. (2011). Comparison of F_{ST} outlier tests for SNP loci under selection. *Molecular Ecology Resources*, **11**, 184–194.

Narum, S. R., Campbell, N. R., Meyer, K. A., Miller, M. R., & Hardy, R. W. (2013). Thermal adaptation and acclimation of ectotherms from differing aquatic climates. *Molecular Ecology*, **22**(11), 3090–3097.

Nei, M. (1972). Genetic distance between populations. *The American Naturalist*, **106**(949), 283–292.

Nomura T. (2008). Estimation of effective number of breeders from molecular coancestry of single cohort sample. *Evolutionary Applications* **1** (3), 462–474.

Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**(3), 375–402.

Orr, H. A. (1998). The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution*, **52**(4), 935-949.

Orsini, L., Marshall, H., Cuenca Cambronero, M., Chaturvedi, A., Thomas, K. W., Pfrender, M. E. *et al.* (2016). Temporal genetic stability in natural populations of the waterflea *Daphnia magna* in response to strong selection pressure. *Molecular Ecology*, **25**(24), 6024–6038.

Paris, J. R., King, R. A., & Stevens, J. R. (2015). Human mining activity across the ages determines the genetic structure of modern brown trout (*Salmo trutta* L.) populations. *Evolutionary Applications*, **8**(6), 573–585.

Paris, J. R., Stevens, J. R., & Catchen, J. M. (2017). Lost in parameter space: a road map for stacks. *Methods in Ecology and Evolution*. DOI 10.1111/2041-210X.12775.

Pavey, S. A., Gaudin, J., Normandeau, E., Dionne, M., Castonguay, M., Audet, C., & Bernatchez, L. (2015). RAD sequencing highlights polygenic discrimination of habitat ecotypes in the panmictic American eel. *Current Biology*, **25**(12), 1666–1671.

Pegg, A. E. (2006). Regulation of ornithine decarboxylase. *The Journal of Biological Chemistry*, **281**(21), 14529–32.

Perreault-Payette, A., Muir, A. M., Goetz, F., Perrier, C., Normandeau, E., Sirois, P., & Bernatchez, L. (2017). Investigating the extent of parallelism in morphological and genomic divergence among lake trout ecotypes in Lake Superior. *Molecular Ecology*, **26**(6), 1477–1497.

Pinto, P. I., Matsumura, H., Thorne, M. A., Power, D. M., Terauchi, R., Reinhardt, R., & Canário, A. V. (2010). Gill transcriptome response to changes in environmental calcium in the green spotted puffer fish. *BMC Genomics*, **11**, 476. **DOI:** 10.1186/1471-2164-11-476.

Plath, M., Hauswaldt, J. S., Moll, K., Tobler, M., García de León, F. J., Schlupp, I., & Tiedemann, R. (2006). Local adaptation and pronounced genetic differentiation in an extremophile fish, *Poecilia mexicana*, inhabiting a Mexican cave with toxic hydrogen sulphide. *Molecular Ecology*, **16**(5), 967–976.

Polyakov, A., Beharav, A., Avivi, A., & Nevo, E. (2004). Mammalian microevolution in action: adaptive edaphic genomic divergence in blind subterranean mole-rats. *Proceedings. Biological Sciences*, (Suppl 4), S156–9.

Poulin, R., & Pegg, A.E. (1990) Regulation of ornithine decarboxylase expression by anisosmotic shock in alpha-difluoromethylornithine-resistant L1210 cells. *J Biol Chem*, **265**(7):4025–4032.

Pritchard, J. K., & Di Rienzo, A. (2010). Adaption – not by sweeps alone. *Nature Reviews Genetics*, **11**(10), 665–667.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, **155**(2), 945–959.

Ravinet, M., Westram, A., Johannesson, K., Butlin, R., André, C., & Panova, M. (2016). Shared and nonshared genomic divergence in parallel ecotypes of *Littorina saxatilis* at a local scale. *Molecular Ecology*, **25**(1), 287–305.

Raymond, M., Rousset, F. (1995) Genepop, Version 1.2: population genetics software for exact tests and ecumenicism. *Journal of Heredity,* **86**, 248–249.

Reid, N. M., Proestou, D. A., Clark, B. W., Warren, W. C., Colbourne, J. K., Shaw, J. R., *et al.* (2016). The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. *Science*, **354**(6317), 1305-1308.

Reznick, D., & Ghalambor, C.K. (2001). The population ecology of contemporary adaptations: what empirical studies reveal about the conditions that promote adaptive evolution. *Genetica*, **112/113**(1), 183–198.

Rotchell, J. M., Lee, J. S., Chipman, J. K., & Ostrander, G. K. (2001). Structure, expression and activation of fish ras genes. *Aquatic Toxicology,* **55**(1-2), 1–21.

Salisbury, S. J., McCracken, G. R., Keefe, D., Perry, R., & Ruzzante, D. E. (2016). A portrait of a sucker using landscape genetics: how colonization and life history undermine the idealized dendritic metapopulation. *Molecular Ecology*, **25**(17), 4126–4145.

Santos, E. M., Hamilton, P. B., Coe, T. S., Ball, J. S., Cook, A. C., Katsiadaki, I., & Tyler, C. R. (2013). Population bottlenecks, genetic diversity and

breeding ability of the three-spined stickleback (*Gasterosteus aculeatus*) from three polluted English Rivers. *Aquatic Toxicology*, **142**, 264–271.

Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics*, **14**(11), 807–820.

Schlamp, F., van der Made, J., Stambler, R., Chesebrough, L., Boyko, A. R., & Messer, P. W. (2016). Evaluating the performance of selection scans to detect selective sweeps in domestic dogs. *Molecular Ecology*, **25**(1), 342–356.

Schubbert, S., Shannon, K., & Bollag, G. (2007). Hyperactive Ras in developmental disorders and cancer. *Nature Reviews Cancer*, **7**(4), 295–308.

Sgrò, C. M., Lowe, A. J., & Hoffmann, A. A. (2011). Building evolutionary resilience for conserving biodiversity under climate change. *Evolutionary Applications*, **4**(2), 326–337.

Shin, H.-J., Park, K.-K., Lee, B.-H., Moon, C.-K., & Lee, M.-O. (2003). Identification of genes that are induced after cadmium exposure by suppression subtractive hybridization. *Toxicology*, **191**(2-3), 121–31.

Sit, S.-T., & Manser, E. (2011). Rho GTPases and their role in organizing the actin cytoskeleton. *Journal of Cell Science*, **124**(5), 679-83.

Smith, T., & Bernatchez, L. (2008). Evolutionary change in human-altered environments. *Molecular Ecology*, **17**(1), 1-8.

Smukowski, C. S., & Noor, M. A. F. (2011). Recombination rate variation in closely related species. *Heredity*, **107**(6), 496–508.

Steffen, W., Grinevald, J., Crutzen, P., & McNeill, J. (2011). The Anthropocene: conceptual and historical perspectives. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **369**(1938), 842–867.

Stockwell, C. A., Hendry, A. P., & Kinnison, M. T. (2003). Contemporary evolution meets conservation biology. *Trends in Ecology & Evolution*, **18**(2), 94–101.

Storz, J. F. (2005). Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, **14**(3), 671–688.

Taylor, E. B. (1991). A review of local adaptation in Salmonidae, with particular reference to Pacific and Atlantic salmon. *Aquaculture*, **98**(1-3), 185–207.

Tétard, S., Feunteun, E., Bultel, E., Gadais, R., Bégout, M.-L., Trancart, T., & Lasne, E. (2016). Poor oxic conditions in a large estuary reduce connectivity from marine to freshwater habitats of a diadromous fish. *Estuarine, Coastal and Shelf Science*, **169**, 216–226.

Thévenod, F. (2010). Catch me if you can! Novel aspects of cadmium transport in mammalian cells. *BioMetals*, **23**(5), 857–875.

Tiffin, P., & Ross-Ibarra, J. (2014). Advances and limits of using population genetics to understand local adaptation. *Trends in Ecology & Evolution*, **29**(12), 673–680.

Uren Webster, T. M., Bury, N., van Aerle, R., & Santos, E. M. (2013). Global transcriptome profiling reveals molecular mechanisms of metal tolerance in a chronically exposed wild population of brown trout. *Environmental Science & Technology*, **47**(15), 8869–77.

Valladares, F., Matesanz, S., Guilhaumon, F., Araújo, M. B., Balaguer, L., Benito-Garzón, M., *et al.* (2014). The effects of phenotypic plasticity and local adaptation on forecasts of species range shifts under climate change. *Ecology Letters*, **17**(11), 1351–1364.

van Oosterhout, C., Joyce, D. A., Cummings, S. M., Blais, J., Barson, N. J., Ramnarine, I. W., *et al.* (2006). Balancing selection, random genetic drift, and genetic variation at the major histocompatibility complex in two wild populations of guppies (*Poecilia reticulata*). *Evolution*, *60*(12), 2562-74.

Vera-Escalona, I., Habit, E., & Ruzzante, D. E. (2015). Echoes of a distant time: effects of historical processes on contemporary genetic patterns in *Galaxias platei* in Patagonia. *Molecular Ecology*, **24**(16), 4112–4128.

Vitousek, P., Mooney, H., & Lubchenco, J. (1997). Human domination of Earth's ecosystems. *Science*, **112**(31), 9511-7.

Wagner, D. N., Baris, T. Z., Dayan, D. I., Du, X., Oleksiak, M. F., & Crawford, D. L. (2017). Fine-scale genetic structure due to adaptive divergence among microhabitats. *Heredity*, DOI: 10.1038/hdy.2017.6.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the analysis of population structure. *Evolution*, **38**(6),1358-1370.

Westley, P. A. H., Ward, E. J., & Fleming, I. A. (2012). Fine-scale local adaptation in an invasive freshwater fish has evolved in contemporary time. *Proceedings of the Royal Society B: Biological Sciences*, **280**(1751), 20122327.

Whitehead, A., Clark, B. W., Reid, N. M., Hahn, M. E., & Nacci, D. (2017). When evolution is the solution to pollution: Key principles, and lessons from rapid repeated adaptation of killifish (*Fundulus heteroclitus*) populations. *Evolutionary Applications*. DOI: 10.1111/eva.12470.

Whitehead, A., Roach, J. L., Zhang, S., & Galvez, F. (2011). Genomic mechanisms of evolved physiological plasticity in killifish distributed along an environmental salinity gradient. *Proceedings of the National Academy of Sciences of the United States of America*, **108**(15), 6193–8.

Whitehead, A., Roach, J. L., Zhang, S., & Galvez, F. (2012). Salinity- and population-dependent genome regulatory response during osmotic acclimation in the killifish (*Fundulus heteroclitus*) gill. *Journal of Experimental Biology*, **215**(8), 1293-1305.

Williams, G.C. (1996). *Adaptation and natural selection : a critique of some current evolutionary thought*. New Jersey, USA: Princeton University Press.

Wu, S. M., Tsai, P. J., Chou, M. Y., & Wang, W.-D. (2013). Effects of maternal cadmium exposure on female reproductive functions, gamete quality, and offspring development in zebrafish (*Danio rerio*). *Archives of Environmental Contamination and Toxicology*, **65**(3), 521–536.

Wu, T. D., & Nacu, S. (2010). Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics*, **26**(7), 873–81.

Zheng, R., Ghirlando, R., Lee, M. S., Mizuuchi, K., Krause, M., & Craigie, R. (2000). Barrier-to-autointegration factor (BAF) bridges DNA in a discrete, higher-order nucleoprotein complex. *Proceedings of the National Academy of Sciences of the United States of America*, **97**(16), 8997–9002.

Zhu, L., Qu, K., Xia, B., Sun, X., & Chen, B. (2016). Transcriptomic response to water accommodated fraction of crude oil exposure in the gill of Japanese flounder, *Paralichthys olivaceus*. *Marine Pollution Bulletin*, **106**(1), 283–291.

Supporting Information

Rapid and repeated local adaptation of brown trout populations to unique cocktails of metal contaminants

This supporting information contains:

Table S1. Details of the analysis of Latent Factor Mixed Models (LFMM).

Figure S1. Histograms of *p*-values for the combined z-scores generated using LFMM

Table S2. Pairwise F_{ST} values between the five populations for microsatellite data and RADseq data

Table S3. Outlier loci identified by F_{ST} outlier methods (Lositan & Bayescan) and LFMM.

Table S4. Gene ontology of all outlier loci.

Table S1. Details of the analysis of Latent Factor Mixed Models (LFMM).

	Arsenic (As)	***Cadmium (Cd)***	***Copper (Cu)***	***Iron (Fe)***	***Nickel (Ni)***	***Zinc (Zn)***
Expected FDR	0.05	0.05	0.05	0.05	0.05	0.05
Estimated FDR	NaN	NaN	NaN	NaN	NaN	NaN
Lambda	0.1395472	0.1372115	0.136778	0.1371841	0.138525	0.138525
True Positives	0	0	0	0	0	0
Adjusted Lambda	0.08	0.08	0.08	0.08	0.08	0.08
Estimated FDR	0.103448276	0.125	0.125	0.125	0.12	0.12
Candidates	29	24	24	24	25	25

Results include False Discovery rate (FDR), genomic inflation estimated as λ, adjusted λ, new estimated FDR, and number of positive candidates as under selection. Results are shown for each one of the metal contaminants tested: arsenic (As), cadmium (Cd), copper (Cu), iron (Fe), nickel (Ni) and zinc (Zn).

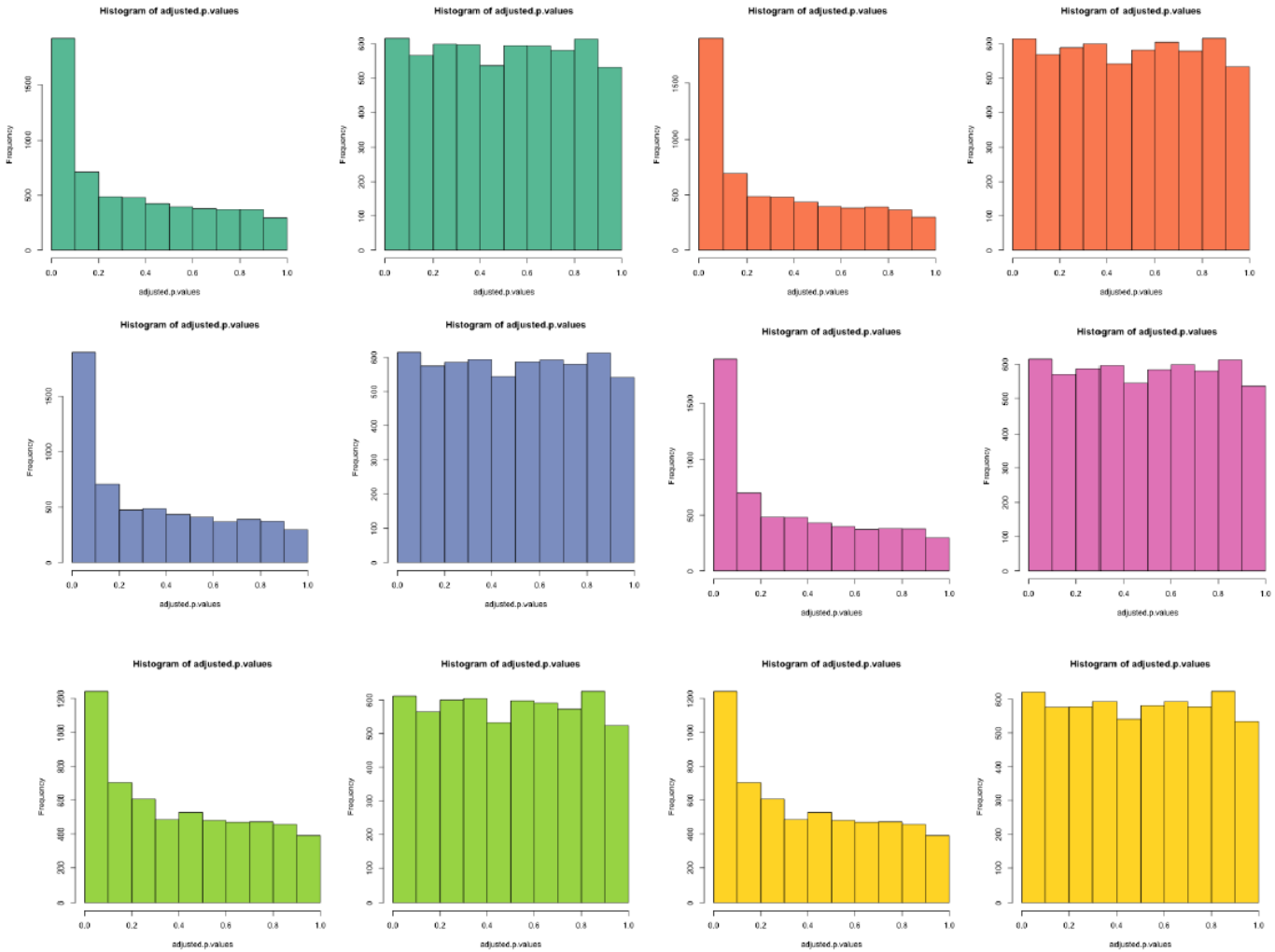


Figure S1. Histograms of *p*-values for the combined z-scores generated using LFMM for each metal pollutant: arsenic (turquoise); cadmium (orange); copper (blue); iron (pink), nickel (green) and zinc (yellow). Each histogram was adjusted by using λ 0.08 (as above) so that each histogram looks flat with a peak close to zero, indicating that the *p*-values are correct (*i.e.*, are drawn from a uniform distribution under the null-hypothesis).

Table S2. Pairwise F_{ST} values between the five populations: Camel (clean); Carrick Roads (clean); Gannel (clean); Hayle (metal) and Red River (metal) for (A) the RADseq data analysed here and (B) the microsatellite data adapted from Paris *et al.* (2015). Clean-river populations are denoted by * and metal-impacted populations are denoted by †

(A)

	Camel *	Carrick Roads *	Gannel *	Hayle †	Red River †
Camel *		0.029	0.039	0.112	0.055
Carrick Roads *	0.001		0.058	0.125	0.069
Gannel *	0.001	0.001		0.112	0.067
Hayle †	0.001	0.001	0.001		0.094
Red River †	0.001	0.001	0.001	0.001	

(B)

	Camel *	Carrick Roads *	Gannel *	Hayle †	Red River †
Camel *		0.024	0.032	0.109	0.058
Carrick Roads *	0.001		0.038	0.116	0.062
Gannel *	0.001	0.001		0.106	0.059
Hayle †	0.001	0.001	0.001		0.112
Red River †	0.001	0.001	0.001	0.001	

Table S3. Outlier loci identified by F_{ST} outlier methods (Lositan & Bayescan) and LFMM for (A) All populations compared; (B) Hayle-specific outlier tests and (C) Red River-specific outlier tests. RAD loci highlighted in yellow represent outliers identified in the all population analyses and in a river-specific analysis. * denotes outliers identified by Lositan and Bayescan; † denotes outliers identified by Bayescan and LFMM; ‡ denotes outliers identified by Lositan and LFMM; and § denotes outliers identified uniquely with arsenic.

(A) All population analysis			(B) Hayle-specific analysis		(C) Red River-specific analysis	
Lositan	*Bayescan*	*LFMM*	*Lositan*	*Bayescan*	*Lositan*	*Bayescan*
461	1913	76	527	*11539	479	*11539
*3908	1995	4406	1068	*21668	1413	*14054
*4204	*3908	5914	1185	*38481	1995	*14577
5312	*4204	†10603	1432	*42147	3193	*18205
‡11460	8091	‡11460	1472	*46160	7273	*23716
*11539	†10603	14103	1543	*55120	10088	*25067
12963	*11539	14566	1715	*71248	10490	*27901
13186	12682	14577	222		10912	*29115
14009	14054	22402	2394		11368	*32833
14659	16508	23204	2522		11539	*60772
*18205	*18205	29611	2546		11834	*63058
20283	21668	33848	3342		12821	
24316	25067	46096	3345		13278	
28972	29115	46426	3661		13542	
30634	32600	47115	3993		13597	
31514	*35593	47908	1498		13694	
31833	*38481	49440	4257		13899	
32224	40829	50293	4266		14054	
*35593	41622	51015	4973		14548	
37628	*42147	51793	5438		14577	
*38481	42448	52678	5944		15318	

Table S3 continued.

(A) All population analysis			(B) Hayle-specific analysis		(C) Red River-specific analysis	
Bayescan	*LFMM*	*Lositan*	*Lositan*	*Bayescan*	*Lositan*	*Bayescan*
38744	*43057	52762	6034		15639	
*42147	*46160	61900	6397		18025	
*43057	*46300	66734	6825		19697	
44846	49455	10701	7049		19792	
45488	50198	§20972	9028		19948	
*46160	50410	§34552	9139		21729	
*46300	*56666	§40921	9655		21893	
47927	57599	§49278	9849		22971	
55120	*59777		9979		23716	
*56666	60772		10652		24542	
56701	63058		10977		25067	
56708	65713		*11539		26227	
57756			12064		26648	
57850			12512		26732	
*59777			12784		26192	
61458			13263		27279	
62959			14009		27901	
63040			14048		28783	
64936			14450		29115	
66688			15713		29319	
71248			15962		29611	

Table S3 continued.

Hayle Lositan				*Red River Lositan*
16714	33869	45488	60163	29925
16868	34777	45685	60283	30034
17183	35207	45940	60602	31004
17566	35271	*46160	61281	31012
17631	35776	46192	61689	32833
17872	35826	46537	61767	37700
18280	36044	47527	62429	38665
18626	36443	47632	63770	39407
19675	36728	47927	65297	39526
19852	37263	48241	65299	42448
20283	37472	48845	65431	45557
20983	37996	48953	65909	46109
*21668	38475	49577	67029	46160
21800	*38481	49605	67080	46703
22204	38744	50175	67224	47283
22339	38989	50500	67237	47502
22370	39782	50652	*71248	47953
23488	40709	50935		48052
24057	40857	50952		4877
24107	40898	51049		48694
24803	41079	51938		53171
25153	41087	52148		54530
25975	*42147	53354		56403
27083	42314	53919		57107
27144	42433	54132		58597
27146	42448	*55120		59087
28865	42854	55246		59882
28972	43349	56287		60772
32387	44718	57756		63058
32931	44846	58560		63719
33508	45049	59246		63861
33568	45269	59643		65241

Table S4. Gene ontology of outlier loci identified: F_{ST} outliers detected using both Lositan and Bayescan, and covarying outlier loci using LFMM. F_{ST} outliers were assigned ontology for the analysis of all populations (All populations); trout populations from the River Hayle (Hayle-specific) and trout populations from the Red River (Red River-specific). For each outlier locus, the RAD locus ID is presented, the chromosome from the Atlantic salmon (*Salmo salar*) and the NCBI accession, the % query cover, the E-value and the % identity from the Blast result. Gene names are given in the final column.

F_{ST} outliers	Locus ID	*S. salar* alignment	Accession	Query cover	E value	Ident	Gene name
All populations	3908	ssa20	NC_027319.1	100%	1.00E-34	97%	clathrin heavy chain 1-like
	4204	ssa10	NC_027309.1	65%	9.00E-17	95%	coiled-coil domain-containing protein 73
	11539	ssa14	NC_027313.1	93%	3.00E-26	97%	antizyme inhibitor 1-like
	18205	ssa11	NC_027310.1	100%	6.00E-38	99%	protein LAP2-like
	35593	ssa18	NC_027317.1	100%	3.00E-36	99%	pro-neuregulin-3, membrane-bound iosform-like
	38481	ssa24	NC_027323.1	100%	6.00E-38	99%	ras-like protein family member 10B
	42147	ssa01	NC_027300.1	100%	6.00E-38	99%	158356 bp at 5' side: zinc finger protein 585A-like / 64152 bp at 3' side: EMILIN-1-like
	43057	ssa28	NC_027327.1	100%	3.00E-36	98%	disintegrin and metalloproteinase domain-containing protein 8-like isoform X4
	46160	ssa25	NC_027324.1	100%	6.00E-38	99%	68406 bp at 5' side: uridine phosphorylase 2-like / 24953 bp at 3' side: plakophilin-4-like
	46300	ssa08	NC_027307.1	100%	1.00E-34	97%	aminoacyl tRNA synthetase complex-interacting multifunctional protein 1
	56666	ssa20	NC_027319.1	100%	1.00E-39	100%	ankyrin repeat domain-containing protein 32
	59777	ssa02	NC_027301.1	100%	3.00E-36	98%	galanin receptor type 1-like

Table S4 continued

F_{ST} outliers	Locus ID	*S. salar* alignment	Accession	Query cover	E value	Ident	Gene name
Red River-specific	11539	ssa14	NC_027313.1	93%	3.00E-26	97%	antizyme inhibitor 1-like
	14054	--	--	--	--	--	--
	14577	ssa24	NC_027323.1	100%	4.00E-40	100%	4013 bp at 5' side: dnaJ homolog subfamily B member 55157 bp at 3' side: transitional endoplasmic reticulum ATPase
	18205	n.a	NC_027310.1	100%	2.00E-36	99%	LAP2-like
	23716	ssa13	NC_027312.1	100%	8.00E-32	95%	rho GTPase-activating protein 32 isoform X1
	25067	ssa14	NC_027313.1	100%	2.00E-37	98%	SUN domain-containing ossification factor-like isoform X2
	27901	ssa17	NC_027316.1	100%	6.00E-33	95%	guanine nucleotide-binding protein G(i) subunit alpha-1-like
	29115	--	--	--	--	--	--
	32833	ssa23	NC_027322.1	100%	1.00E-24	89%	104486 bp at 5' side: receptor-type tyrosine-protein phosphatase F-like isoform X6381428 bp at 3' side: uncharacterized protein LOC106584099
	60772	ssa17	NC_027316.1	100%	6.00E-38	99%	104114 bp at 5' side: zinc finger E-box-binding homeobox 2-like isoform X5137885 bp at 3' side: kynureninase-like
	63058	ssa16	NC_027315.1	56%	4.00E-15	96%	40115 bp at 5' side: fidgetin-like isoform X343354 bp at 3' side: complement C1q-like protein 2

Table S4 continued

LFMM outliers	Locus ID	*S. salar* alignment	**Accession**	**Query cover**	**E value**	**Ident**	
All metals	76	ssa10	NC_027309.1	100%	1.00E-34	97%	6901 bp at 5' side: hydroperoxide isomerase ALOXE3-like / 7440 bp at 3' side: arachidonate 15-lipoxygenase B-like
	4406	ssa28	NC_027327.1	98%	4.00E-30	94%	42567 bp at 5' side: chromobox protein homolog 2-like / 23642 bp at 3' side: complement C1q tumor necrosis factor-related protein 1
	5914	ssa16	NC_027315.1	100%	3.00E-36	98%	phosphatidylinositol 4-kinase alpha-like
	10603	ssa02	NC_027301.1	100%	6.00E-38	99%	mucin-22-like
	11460	ssa01	NC_027300.1	100%	3.00E-36	98%	protein phosphatase 1 regulatory subunit 21-like
	14103	ssa11	NC_027310.1	100%	1.00E-34	97%	glutamine--fructose-6-phosphate aminotransferase [isomerizing]
	14566	ssa01	NC_027300.1	100%	6.00E-38	99%	nesprin-2-like
	14577	ssa24	NC_027323.1	100%	1.00E-39	100%	4013 bp at 5' side: dnaJ homolog subfamily B member 5 / 5157 bp at 3' side: transitional endoplasmic reticulum ATPase
	22402	ssa21	NC_027320.1	100%	6.00E-38	99%	sodium leak channel non-selective protein
	23204	ssa03	NC_027302.1	98%	5.00E-39	100%	E3 ubiquitin-protein ligase RNF213
	29611	ssa01	NC_027300.1	100%	1.00E-34	97%	268082 bp at 5' side: doublecortin domain-containing protein 2-like / 655722 bp at 3' side: radical S-adenosyl methionine domain-containing protein 2-like
	33848	ssa16	NC_027315.1	100%	6.00E-38	99%	45009 bp at 5' side: aurora kinase B-like / 20466 bp at 3' side: alpha-1,3-mannosyl-glycoprotein 4-beta-Nacetylglucosaminyltransferase C-like
	46096	ssa03	NC_027302.1	97%	3.00E-21	91%	10074 bp at 5' side: matrix metalloproteinase-14-like / 6575 bp at 3' side: apoptotic chromatin condensation inducer in the nucleus-like
	46426	ssa14	NC_027313.1	100%	1.00E-34	97%	pro-FMRFamide-related neuropeptide FF

Table S4 continued

LFMM outliers	Locus ID	*S. salar* alignment	Accession	Query cover	E value	Ident	
	47115	ssa12	NC_027311.1	100%	1.00E-34	97%	cadherin-22-like
	47908	--	--	--	--	--	--
	49440	ssa04	NC_027303.1	100%	6.00E-38	99%	6192 bp at 5' side: protocadherin Fat 1-like / 525766 bp at 3' side: FRG1 protein
	50293	--	--	--	--	--	--
	51015	ssa22	NC_027321.1	100%	2.00E-37	99%	collagen alpha-1(II) chain
	51793	ssa14	NC_027313.1	100%	6.00E-38	99%	rho-associated protein kinase 1-like
	52678	ssa14	NC_027313.1	98%	2.00E-27	92%	37339 bp at 5' side: angiopoietin-1-like / 172907 bp at 3' side: syndecan-2-A-like
	52762	ssa08	NC_027307.1	100%	3.00E-36	98%	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase beta
	61900	ssa16	NC_027315.1	100%	1.00E-39	100%	6306 bp at 5' side: ras-related protein Rab-8A / 12151 bp at 3' side: calcium and integrin-binding family member 3
	66734	ssa29	NC_027328.1	100%	1.00E-34	97%	disks large-associated protein 1-like
Arsenic, nickel & zinc	10701	ssa18	NC_027317.1	100%	1.00E-39	100%	RNA-binding protein 4B-like
Arsenic only	20972	ssa05	NC_027304.1	100%	1.00E-34	97%	Calpain-7
	34552	ssa26	NC_027325.1	100%	1.00E-39	100%	homeobox protein orthopedia B-like isoform X1
	40921	ssa11	NC_027310.1	100%	7.00E-28	92%	ALK tyrosine kinase receptor-like
	49278	ssa01	NC_027300.1	100%	3.00E-31	94%	uncharacterized protein C14orf132-like

Chapter V

Brown trout display multi-faceted physiological mechanisms to cope with chronic metal toxicity in a metal-impacted river

Manuscript in preparation

AUTHORS: J.R. Paris[1], M.A. Urbina[2], R.A. King[1], D. Rowe[1], R.W. Wilson[1], E.M. Santos[1], N.R. Bury[3] & J.R. Stevens[1]

[1] Biosciences, College of Life and Environmental Sciences, University of Exeter, United Kingdom

[2] Departamento de Zoología, Facultad de Ciencias Naturales y Oceanográficas, Universidad de Concepción, Concepción, Chile

[3] Faculty of Science, Health and Technology, University of Suffolk, Waterfront Building, Neptune Quay, Ipswich, IP4 1QJ, United Kingdom

J.R. Paris undertook the exposure experiments, conducted the laboratory work and the data analysis, and wrote the manuscript.

M.A. Urbina assisted in conception of the exposure experiments, conducted laboratory work and data analysis, and contributed to writing of the manuscript.

R.A. King assisted with the exposure experiments, undertook laboratory work and improved drafts of the manuscript.

D. Rowe conceived the exposure experimental design and set up, maintained and assisted with the exposure experiments.

R.W. Wilson conceived the exposure experiments and assisted with data analysis.

E.M. Santos conceived the exposure experiments and assisted with data analysis.

N.R. Bury conceived the exposure experiments, contributed to data analysis and improved drafts of the manuscript.

J.R. Stevens conceived the exposure experiments, provided support throughout and improved drafts of the manuscript.

Brown trout display multi-faceted physiological mechanisms to cope with chronic metal toxicity in a metal-impacted river

Manuscript in preparation

AUTHORS Paris JR[1], Urbina MA[2], King RA[1], Rowe D[1], Wilson RW[1], Santos EM[1], Bury NR[3] & Stevens JR[1]

[1] Biosciences, College of Life and Environmental Science, Geoffrey Pope Building, University of Exeter, Exeter, EX4 4QD, United Kingdom

[2] Departamento de Zoología, Facultad de Ciencias Naturales y Oceanográficas, Universidad de Concepción, Concepción, Chile

[3] Faculty of Science, Health and Technology, University of Suffolk, Waterfront Building, Neptune Quay, Ipswich, IP4 1QJ, United Kingdom

ABSTRACT

Metal pollution is known to negatively impact aquatic ecosystems across the globe, yet some fish populations demonstrate processes indicative of acclimation and/or adaptation to chronically polluted environments. Both excesses and deficiencies in essential metals can cause detrimental health impacts, and so mechanisms involved in the homeostasis and detoxification of metals are necessary. Fish can mitigate the potential harm of toxic metals through various physiological processes in metal handling, including reduced uptake, storage and excretion. In this study, we investigated processes of metal tissue accumulation, osmoregulation and ammonia excretion in brown trout (*Salmo trutta* L.) sampled from a chronically metal-impacted river, the River Hayle and a clean reference river, the River Fowey. We assessed these fish in the in-river environment, after a depuration period, and during a 96-hour exposure to a mixture of metals: Cd 0.04 µg/L, Cu 3 µg/L, Ni 10 µg/L, Zn 75 µg/L. Metal-tissue burden analysis showed that the two populations have distinctive tissue accumulation patterns, pointing to different processes underlying the regulation of essential and non-essential metals. Osmoregulation differed between the two populations, with the Fowey fish showing an over-compensatory response to metal exposure. Ammonia excretion rates were highly significantly different between the two populations, even before initiation

of the metal exposure. We conclude that trout occupying the River Hayle display distinct patterns of metal handling, pointing to various mechanisms involved in inherent physiological adaptation.

KEYWORDS: metal pollution, acclimation, adaptation, exposure, cadmium, copper, nickel, zinc, osmoregulation, sodium, potassium, Na^+/K^+-ATPase, ionic mimicry

INTRODUCTION

Anthropogenic pollution affects natural habitats worldwide. Amongst the most prevalent pollutants in the environment are metals (Förstner & Wittmann 1989; Nriagu 1990; Tchounwou *et al.* 2012), for which natural background levels have been significantly augmented by human activity (Boutron *et al.* 1995; Candelone *et al.* 1995; Nriagu 1996). Metal pollution has been shown to impact fish and fish populations in a number of ways including physiological damage (Eroglu *et al.* 2015; Fernandes *et al.* 2013), diminished genetic diversity (Maes *et al.* 2005; Paris *et al.* 2015), altered behaviour (Baatrup 1991; Scott & Sloman 2004), impaired growth, reproduction and fecundity (Canli & Atli 2003; Boyle *et al.* 2008; Crump & Trudeau 2009; Witeska *et al.* 2014; Hoseini *et al.* 2016), and in some cases, fatalities and local extinctions (Van Hoof & Van San 1981; Bradley & Morris 1986; Pain *et al.* 1998). However, some fish populations display a remarkable capacity to withstand chronic metal concentrations without obvious detrimental effect (Hansen *et al.* 2006; Bourret *et al.* 2008; Durrant *et al.* 2011; Paris *et al.* 2015).

Physiological mechanisms such as the induction of metal-binding proteins (Hogstrand & Haux 1991), alterations in energy metabolism (Bougas *et al.* 2016) and adjustments in iono- and osmo-regulation (Atli & Canli 2013) all play important roles in mitigating metal toxicity. However, in response to metal exposure, an organism will make adjustments to these processes, and the extent to which it can withstand the perturbations caused by metals will depend on the inherent plasticity of this physiology - a process termed metal acclimation (McDonald & Wood 1993; McGeer *et al.* 2007). Alternatively, genetic adaptation, by means of selection for metal tolerance traits, may confer significant resistance to metal toxicity in chronically exposed fish populations (Ownby *et al* 2002; Meyer & Di Giulio 2003).

Gills feature a large surface area and are in intimate contact with the water (Hughes 1984) and are therefore essential in maintaining homeostasis between a fish and its environment (Evans 1987). Metals are known to interfere with the gill through various processes, including ionic mimicry of essential ions (Bridges & Zalups 2005) and competitive binding and inhibition of ion transport proteins (Playle *et al.* 1993; Bury *et al.* 1999; Grosell *et al.* 2004; Atli & Canli 2007). Such

processes are known to result in reduced ion uptake at the gill, causing osmoregulatory stress (Wood *et al.* 1996; Grosell *et al.* 2002). Perturbations to gill function are therefore a sensitive indicator of the effects of metals (Laurén & McDonald 1985; Wood 1992; Playle 2004). Measuring potential disruptions in osmoregulation via quantification of net increases or decreases of Na^+ and Cl^- in water and blood plasma, and the activity of ionoregulatory protein channels (Na^+/K^+-ATPase) is a useful means of quantifying this (Chowdhury *et al.* 2016; Atli & Canli 2011).

Patterns of tissue metal accumulation provide an insight into the mechanisms of metal-specific and tissue-specific metal handling. Some metals are essential for life (e.g. Cu, Fe, Zn), but in excess they are toxic (Skidmore 1964; Laurén & McDonald 1989; Dalzell & Macfarlane 1999). Conversely, deficiencies in these metals undermine health (Kawatsu 1972; Spry *et al.* 1988; Kamunde *et al.* 2002a). Other metals, such as Cd and Pb, are non-essential and in excess are detrimental to health (Benoit *et al.* 1976; Weber 1993). Studies quantifying metal accumulation have shown that metals may not readily depurate from tissues, and may also vary seasonally (Gill *et al.* 1992; Dural *et al.* 2007). Different metals exhibit differential uptake and partitioning patterns within organs (Hansen *et al.* 2006; Poleksic *et al.* 2010; Monna *et al.* 2011), and processes of sub-cellular partitioning within these organs are thought to represent a significant mechanism attributable with acclimation to chronic metal exposure (Kraemer *et al.* 2005; Giguère *et al.* 2006).

Given the balance between necessity and toxicity, various biological processes must exist in fish to maintain metal equipoise; these include uptake and export, as well as partitioning into organelles and detoxification. Through these means, fish can control the homeostatic concentrations of intracellular metal levels and ultimately, avoid negative toxic effects. Sub-cellular partitioning, whereby trace metals bind to physiologically sensitive target molecules, for example, to a metal-sensitive fraction (MSF) or a metal-detoxified fraction (MDF) represents an important process in reducing the toxic impact of metals (see review by Mason & Jenkins 1995). In particular, one strategy in detoxification involves storage of excess metal in metal-rich granules (MRGs) or lysosomes within tissues, through subcellular partitioning into the MDF (Lanno *et al.* 1987; Klerks

& Bartholomew 1991; Lapointe & Couture 2009). Thus, tissue metal concentrations may be high, but the metals within the tissue will not be toxic (e.g. Wang *et al.* 2011; Gimbert *et al.* 2016).

The metal resources of southwest England have been exploited for thousands of years, resulting in several polluted rivers across the region (Pirrie *et al.* 1997, 1999, 2002, 2003). The River Hayle, situated on the northern coast of Cornwall in southwest England, represents an extreme case of a metal-impacted river, for which the effects of metal toxicity on aquatic life have been extensively studied (Brown 1976, 1977; Butler *et al.* 1980; Foster 1982, Gower *et al.* 1994; Khan *et al.* 2011). Metal pollution in the River Hayle can reach extremely high levels (As 99 µg/L; Cd 2.58 µg/L; Cu 417 µg/L; Zn 2512 µg/L), and these concentrations periodically peak above the LC50 values for salmonids (Chakoumakos *et al.* 1979; Todd *et al.* 2009) and specifically, brown trout (*Salmo trutta* L.) (Everall *et al.* 1989; Brinkman & Hansen 2007; Brinkman & Woodling 2014).

Despite the potential ecotoxicological effects of this polluted environment, brown trout populations are known to reside in the River Hayle, and show extremely elevated concentrations of metals in their tissues (Uren-Webster *et al.* 2013). Studies of the Hayle trout have documented various patterns and processes, which, considered together, suggest that these fish are metal-tolerant. These include marked differences in neutral genetic patterns (Paris *et al.* 2015), and pronounced regulation of gene expression, including up-regulation of metallothioneins (MTs), other metal-binding proteins and ion transporters (Uren-Webster *et al.* 2013). River water from the Hayle was shown to cause ionoregulatory disturbances in metal-naïve fish (Durrant *et al.* 2011) and upregulated gene expression in the gills of Hayle fish suggests a potential role of ionoregulatory adaptations (Uren-Webster *et al.* 2013). To date, however, a comprehensive laboratory exposure experiment to understand the physiological mechanisms allowing these fish to survive in these metal-laden waters has yet to be undertaken.

In the present study, we conducted a 96-hour exposure of brown trout to a metal mixture representing the most common metal contaminants of the River Hayle: Cd, Cu, Ni and Zn (Minghetti *et al.* 2014). We investigated the

physiological mechanisms of metal tolerance in a population of metal-impacted brown trout from the River Hayle, in comparison to the metal-handling physiology of fish from a nearby control population, the River Fowey. Importantly, this controlled environmental experiment allowed us to examine the relative contribution of a number of physiological mechanisms, potentially acting in synergy, which confer tolerance to metal toxicity in a chronically exposed wild fish population. Ultimately, this physiological understanding assists our knowledge of how natural fish populations can thrive in sub-optimal environments.

MATERIALS AND METHODS

Sampling sites and fish maintenance

Brown trout were sampled from the River Hayle (50°90'24.7"N, 5°25'58.8"W) as the metal-impacted population, and from the River Fowey (50°32'03.5"N 4°36'39.6"W) as the control reference population. Trout from the River Fowey were chosen as a control population as the river experiences low levels of contemporary metal contamination (Table 1). Trout from each river were sampled one week apart in September 2015.

Wild brown trout (0+ fry) were caught by electrofishing under Environment Agency authorisation and transported back to a laboratory at the Aquatic Resources Centre (ARC), University of Exeter. The experimental set-up room was temperature controlled at 15°C and a natural day-night light cycle was used. For both populations, and for each treatment group (hereafter referred to as 'control' and 'exposed'), fish were kept in cohorts of nine to 11 individuals in 39-litre tanks. The fish were fed daily on a diet of frozen bloodworm and live daphnia for 7 days, until 24 hours pre-exposure when feeding was ceased.

Table 1. *Water chemistry features of the native rivers of the two brown trout populations used in this study: Hayle* (metal-impacted) *and Fowey* (clean).

	Fowey	*Hayle*
Site location	50°32'03.5"N 4°36'39.6"W	50°90'24.7"N, 5°25'58.8"W
pH	7.3 ± 0.04	7.38 ± 0.03
Cond @ 20C uS/cm	90.5 ± 3.50	316.78 ± 7.60
Cond @ 25C uS/cm	99.57 ± 1.92	325.78 ± 5.80
Hardness mg/l	18.36 ± 0.49	100.41 ± 2.10
Alky pH 4_5 mg/l	9.19 ± 0.63	42.30 ± 0.06
Chloride (Cl) mg/l	15.5 ± 0.00	44.46 ± 0.90
Sodium (Na) mg/l	9.18 ± 0.00	24.70 ± 1.20
Potassium (K) mg/l	1.71 ± 0.20	4.43 ± 1.80
Calcium (Ca) mg/l	4.17 ± 0.12	20.30 ± 3.20
Arsenic (As) - µg/L	4.57 ± 0.00	14.17 ± 3.39
Cadmium (Cd) - µg/L	0.09 ± 0.00	1.39 ± 0.03
Copper (Cu) - µg/L	4.3 ± 0.00	48.92 ± 2.98
Iron (Fe) - µg/L	64.02 ± 10.90	437.93 ± 151.72
Nickel (Ni) - µg/L	0.63 ± 0.00	26.76 ± 0.87
Zinc (Zn) -µg/L	13.81 ± 0.71	630.80 ± 12.70

Values represent averages of the past five years of metal contamination data obtained from the fish sampling location. These data were provided by the Environment Agency.

Water chemistry

All fish were maintained in a flow-through system of synthetic freshwater, comprised of a chemistry adapted from Minghetti *et al.* (2014) using salts of the following concentrations: NaCl 22.9 mg/L; $KHCO_3$ 12.8 mg/L; $MgSO_4$ 123.2 mg/L and $Ca(NO_3)_2$ 100.4 mg/L. Salts were made up at a 50x stock concentration that was dosed into constant-flowing reverse osmosis (RO) water using a conductivity pump set to dose between 265-270 µS/cm. Water was supplied to fish tanks via a header tank system at a flow rate of 100ml/min. Daily measurements of pH and conductivity of the header tank were monitored throughout the experiment and samples of water were taken from control and exposed tanks for metal-burden analysis.

Exposure and sampling

Fish were sampled on immediate return to the laboratory (termed 'in-river') and the remaining fish were maintained in control water for a depuration phase of seven days (Figure 1). To determine whether metals had depurated from the fish tissues, we sampled at a time point after depuration but prior to the exposure (termed 'depuration'). Both Hayle and Fowey fish were then subjected to a 96-hour exposure, whereby populations were split into control and exposed cohorts. The sub lethal exposure comprised a metal mixture of: Cd 0.4 µg/L; Cu 3 µg/L; nNi 10 µg/L and Zn 75 µg/L. Each metal was quantified to target concentrations using the following salts: $Cd(NO_3)_2 \cdot 4H_2O$ (308.48 g/mol); $CuSO_4 \cdot 5H_2O$ (249.69 g/mol); $NiCl_2 \cdot 6H_2O$ (237.69 g/mol); $ZnSO_4 \cdot 7H_2O$ (287.58 g/mol). All salts used were Sigma-Aldrich, >99.997% trace metal basis.

Sampling was conducted at 24 hours ('24 hours control' and '24 hours exposed') and 96 hours ('96 hours control' and '96 hours exposed') (Figure 1). Fish were humanely killed via overdose of benzocaine (0.5 g L^{-1}), followed by destruction of the brain, under UK Home Office regulations. To account for any sex-related biases in the results, all fish were genotyped using a trout-specific sex marker (King *et al. pers comm*). We measured the morphometric parameters: (i) weight (g) and length (cm); (ii) condition factor (K = weight (g) x 100) / (fork length $(cm)^3$); (iii) hepatosomatic index (HSI = (liver weight (g) / (total weight (g)-liver weight (g)) x 100) and (iv) haematocrit (HCT = total blood (mm)/ red blood cells (mm)). Gill, gut, kidney, liver and muscle tissues were

sampled for metal burden analysis and stored at -20°C until analysis.

Metal burden chemistry analysis

Tissues were digested in 500µl of nitric acid (70%, purified by redistillation, ≥99.999% trace metals basis, Sigma Aldrich) for 48 hours at room temperature (~20°C), followed by 0.1% hydrogen peroxide (Fisher; Hydrogen Peroxide 30% (w/v) Extra Pure SLR) for a further 24 hours before digested tissue was diluted 1:10 in ultrapure deionised H_20. Water samples were acidified with 0.1% nitric acid. For within-sample correction, a 10µg/L yttrium (Y) standard was added to each sample. Metal concentrations (quantified as ng/mg) were measured by ICP-MS (E:AN 6100DRC, Perkin-Elmer, Cambridge, UK).

Metal exposure disturbances in ionoregulation and ammonia handling.

In order to assess differences in ionoregulation, osmolality and ammonia excretion between Fowey and Hayle exposed fish, we measured the net fluxes of the key ions implicated in metal-impacted gill responses, sodium (Na^+), chloride (Cl^-), and also measured net fluxes of ammonia (NH_3) (Figure 1). Due to the sampling regime, we were unable to measure Ca^{2+} fluxes, which are also affected by Cd and Zn exposure. We also quantified overall plasma osmolality, concentrations of Na^+ and Cl^- ions in the blood plasma, and the protein activity of Na^+/K^+-ATPase. To obtain these data, individual fish from each population (n=10; Fowey weight 3.44g ± 0.27; Hayle weight 5.64g ± 0.33) were placed in 0.5 litre flux chambers and allowed to acclimate to chamber conditions for 24 hours. Fish were then exposed for 96 hours to the metal mixture comprising Cd, Cu, Ni and Zn at the concentrations outlined above. For Na^+, Cl^- and NH_3 measurements, water samples were obtained from the tank at the start (flux initial -F_i), and at the end (flux final - F_f), of each flux period as outlined in Table 2.

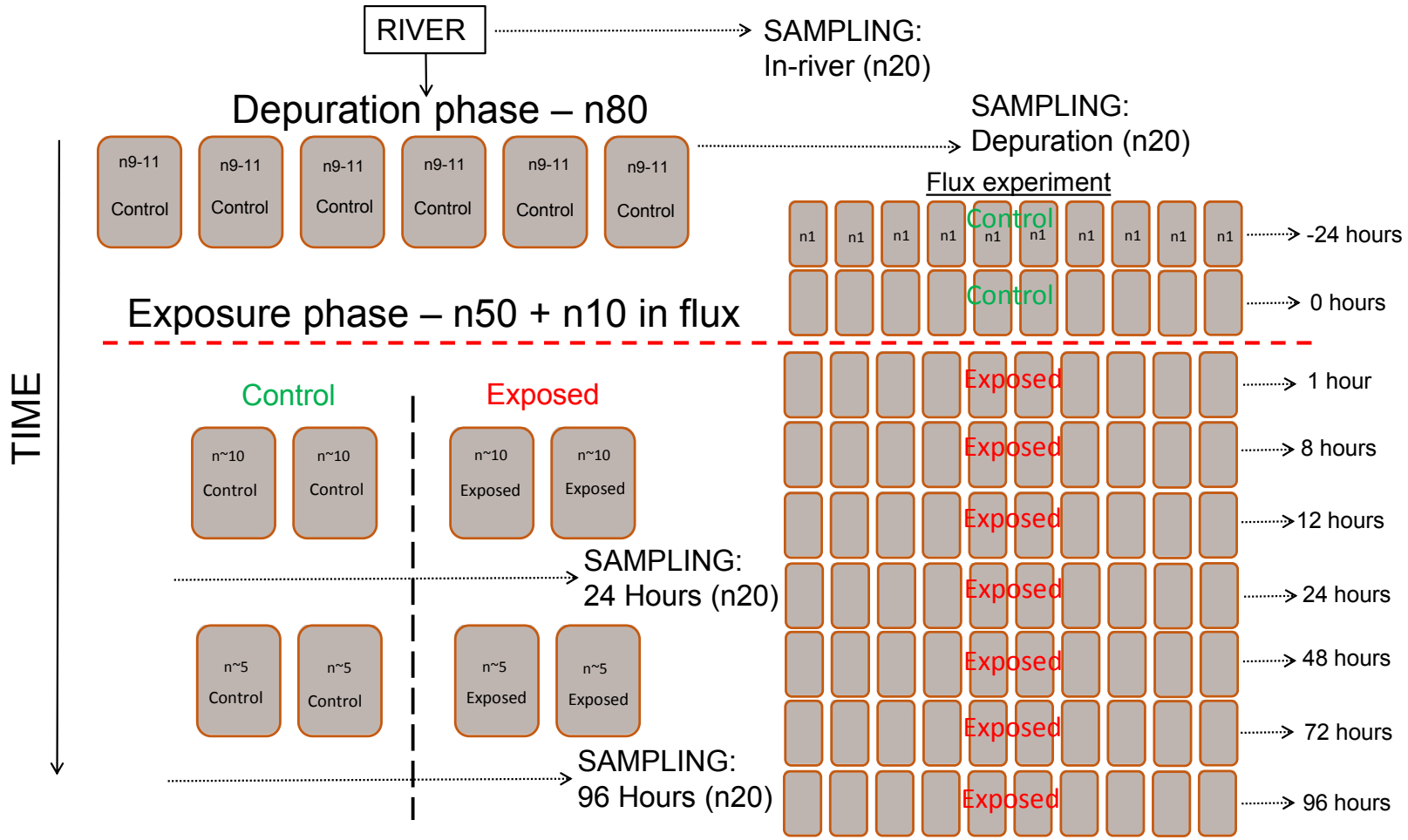


Figure 1. Experimental design of the exposure experiment. Approximately 100 fish were sampled from each river (Hayle and Fowey). Sampling occurred on immediate return to the laboratory (n20). Remaining fish were maintained in control water for 7 days. For the main experiment (left), sampling occurred after the depuration period (n20) and remaining fish were split into control and exposed cohorts (n10 per cohort). Sampling occurred at 24 hours and 96 hours. For the flux experiment (right), 10 fish were placed in control water for 24 hours before the start of the exposure phase, and all fish were then exposed. Throughout the flux experiment, water samples were obtained at the intervals indicated (also displayed in Table 2).

Table 2. *Flush and flux time points across the 96-hour exposure*.

Time point	Nominal time point	Flush/Flux	Duration (hours)	Experimental Conditions
-48 to -24 hours		Flush	24	Control
-24 to -21 hours	-24 hours	Flux	3	Control
-21 to -3 hours		Flush	18	Control
-3 to 0 hours	0 hours	Flux	3	Exposure
0-1 hours		Flush	1	Exposure
1-4 hours	1 hour	Flux	3	Exposure
4-5 hours		Flush	1	Exposure
5-8 hours	8 hours	Flux	3	Exposure
8-9 hours		Flush	1	Exposure
9-12 hours	12 hours	Flux	3	Exposure
12-24 hours		Flush	12	Exposure
24-30 hours	24 hours	Flux	6	Exposure
30-48 hours		Flush	18	Exposure
48-54 hours	48 hours	Flux	6	Exposure
54-72 hours		Flush	18	Exposure
72-78 hours	72 hours	Flux	6	Exposure
78-96 hours		Flush	18	Exposure
96-102 hours	96 hours	Flux	6	Exposure

Each flux initial (F_i), and flux final (F_f) time point comprised a total flux period of either three or six hours. Nominal time points are used to represent each flux period. A total of nine flux measurements were taken, including time points reflecting pre-exposure conditions (-24 hours and 0 hours).

Concentrations of Na^+ and Cl^- of the water samples were determined by ion chromatography (Dionex ICS-1100; Thermo Fisher Scientific, Massachusetts, USA). Fluxes were measured using the equation of Wood (1992) where the *net fluxes* of each ion (Na^+ and Cl^-) are calculated based on appearance or disappearance from the water. In order to determine the blood plasma ion concentrations of Na^+ and Cl^-, plasma was diluted 201-fold in ultrapure deionised H_20 and concentrations of each ion were measured by ion chromatography (Dionex ICS-1100). Plasma osmolality was determined using 10µl of centrifuged blood in a vapour pressure osmometer (Wescor 5520; Wescor Inc., South Logan, UT, USA). NH_3 was quantified spectrophotometrically using the indophenol method (Ivančić & Degobbis 1984) and net fluxes were quantified using the calculation of Wood (1992). Na^+/K^+-ATPase activity of the gills was quantified following a modified protocol as outlined in Urbina & Glover (2015); results were expressed as µmol of ADP per mg $protein^{-1}$ h^{-1}.

Statistical treatment

Comparisons of condition factor parameters were determined using a one-way ANOVA, or a t-test with multiple comparison correction (FDR <0.05). We used a two-way ANOVA to assess metal tissue burden across populations and time points, treating population as the first factor, and exposure time as the second factor. Flux data for the ions Na^+, Cl^-, as well as ammonia (NH_3) were also treated with a two-way ANOVA. For all two-way ANOVA analyses, any significant differences were further explored with pairwise multiple comparison procedures using the Holm-Sidak method. A p-value of <0.05 was accepted for testing significance.

RESULTS

Water and exposure chemistry

Water chemistry of the synthetic freshwater remained consistent throughout depuration and exposure for both populations: Fowey pH 7.3 ± 0.01; 15.5 °C ± 0.02, 269 µS/cm ± 0.53; Hayle pH 7.2 ± 0.01; 15.5 °C ± 0.02, 269 µS/cm ± 0.49 (Table S1). There were statistically significant differences between the ambient metal levels of the control and exposed groups for both populations (Table S2). Fowey Cd control 0.02 µg/L ± 0.01, Cd exposed 0.106 µg/L ± 0.01 (p=0.0016);

Cu control 0.64 ± 0.12 µg/L , Cu exposed 7.82 ± 0.39 µg/L ($p<0.0001$); Ni control 0.75 ± 0.23 µg/L, Ni exposed 11.48 ± 0.16 µg/L ($p<0.0001$); Zn control 2.02 ± 0.59 µg/L, Zn exposed 140.33 ± 18.72 µg/L ($p<0.0001$). Hayle Cd control 0.04 ± 0.01 µg/L, Cd exposed 0.09 ± 0.02 µg/L ($p<0.0001$); Cu control 0.83 ± 0.15 µg/L, Cu exposed 6.49 ± 0.1 µg/L, Ni control 1.73 ± 1.07 µg/L, Ni exposed 13.14 ± 0.76 µg/L ($p<0.0001$), Zn control 2.86 ± 0.7 µg/L, Zn exposed 120.56 ± 4.0 µg/L ($p<0.0001$). There were slight differences between the mean concentrations of metals in the exposure for each population: Fowey Cd 0.1 ± 0.01 µg/L, Hayle Cd 0.1 ± 0.02 µg/L ($p=0.42$); Fowey Cu 8 ±0.39 µg/L, Hayle Cu 6 ± 0.1 µg/L ($p=0.03$); Fowey Ni 11 ± 0.16 µg/L, Hayle Ni 13 ± 0.76 µg/L ($p=0.099$); Fowey Zn 140 ±18.72 µg/L, Hayle Zn 121 ± 4.0 µg/L ($p=0.360$).

Condition factor parameters

All fish were observed feeding; there were no fish mortalities and no negative health effects were detected. The Fowey and Hayle fish had approximately even sex ratios (1:1 and 1:1.18, respectively), but this varied between control and exposure cohorts (Table S3). Sex ratio differences did not change the outcome of the data. Despite the trout from each population being categorised as the same age class at sampling, the Fowey fish were smaller than the Hayle fish in respect to weight (Fowey average 2.61 ± 1.04 g; Hayle average 4.49 ± 1.40 g; $p<0.001$) and fork-length (Fowey average, 6.24 ± 0.87 cm; Hayle average 7.60 ± 0.93 cm; $p<0.001$). Both populations of fish showed a general decline in condition factor *K* from the in-river environment compared to laboratory conditions ($F=4.13$; $P=0.002$); however, this did not alter with the metal exposure.

Across all time points, the hepatosomatic index (HSI) was significantly lower in the Hayle fish compared to the Fowey fish ($F=24.003$; $p<0.001$; Table 3), but there were no exposure-specific responses observed in either population. The Hayle fish showed no changes in HCT across any of the time points during the experiment, with no observed shifts from in-river to depuration or on re-exposure to metals (Table 3). On the other hand, Fowey exposed fish showed significant increases in HCT in comparison to Fowey controls at 96 hours ($F=4.404$; $p<0.001$); the largest of these increases were observed between Fowey fish in-river HCT compared to 96-hours exposed HCT ($t=5.121$; $p<0.001$). Furthermore, in comparison to Hayle exposed fish, Fowey fish

showed a significant increase in HCT at both 24-hours exposed ($t=3.940$; $p<0.001$) and 96-hours exposed ($t=3.808$; $p<0.001$). HCT of Hayle fish across all time points, and Fowey fish maintained in control conditions, showed HCT values typical of freshwater salmonids: 31-34% (McWilliams 1980; Ruane *et al.* 1999; Sloman *et al.* 2000).

Table 3. *Averages and standard error for condition factor parameters of the Fowey and Hayle populations: condition Factor (K); hepatosomatic index (HSI); haematocrit (HCT) and plasma osmolality.*

	Time	*Fowey*	*Hayle*
Condition factor *K*	In-river	1.12 ± 0.032	1.07 ± 0.027
	Depuration	1.02 ± 0.025	1.01 ± 0.014
	24 hours control	1.02 ± 0.022	0.97 ± 0.064
	24 hours exposed	0.99 ± 0.030	0.96 ± 0.067
	96 hours control	0.97 ± 0.013	0.99 ± 0.024
	96 hours exposed	0.97 ± 0.011 †	0.97 ± 0.019 †
HSI	In-river	0.93 ± 0.124 *	0.75 ± 0.044*
	Depuration	1.77 ± 0.319 *	0.84 ± 0.080 *
	24 hours control	1.01 ± 0.126 *	0.81 ± 0.177 *
	24 hours exposed	1.20 ± 0.189 *	0.78 ± 0.083 *
	96 hours control	0.96 ± 0.117 *	0.62 ± 0.073 *
	96 hours exposed	1.36 ± 0.184 *	0.77 ± 0.116 *
HCT *mm*	In-river	32.44 ± 1.849	34.24 ± 2.078
	Depuration	36.2 ± 1.519	33.87 ± 1.121
	24 hours control	35.69 ± 1.644	36.75 ± 1.235
	24 hours exposed	40.52 ± 2.173 *	31.67 ± 1.283 *
	96 hours control	35.33 ± 1.611	37.68 ± 1.289
	96 hours exposed	44.39 ± 1.468 †*	36.3 ± 0.930 *
Osmolality *mmol/kg*	In-river	306.8 ± 9.8 *	279.92 ± 5.47 *
	Depuration	301.27 ± 4.83 *	317 ± 7.0 †*
	24 hours control	305.7 ± 3.38	292 ± 5.16
	24 hours exposed	304.67 ± 3.72	299.67 ± 4.0
	96 hours control	299.86 ± 5.9	308.21 ± 1.7
	96 hours exposed	295.73 ± 4.7 *	304.57 ± 1.8 *

Each parameter was measured throughout the experiment at the following time-points: in-river; depuration; 24 hours (control and exposed) and 96 hours (control and exposed). * represents significance between populations and † represents significant changes within a population.

Disruptions and adjustments in osmoregulation

There was a significant interaction between time and population for measurements of J_{net} Na^+ (F=2.43; p=0.017; Figure 2; Table S4). Significant differences were observed between the Fowey and Hayle fish at 1 hour (t=3.1; p=0.002) and 96 hours (t= 2.26; p=0.025). The significant results at 1 hour were driven by a relative net increase in Na^+ fluxes of Fowey fish, which was also significantly higher than measurements of Fowey fish analysed at -24 hours (p<0.001), 8 hours (p=0.016), 12 hours (p<0.001) and 96 hours (p=0.017). From 8 hours onwards, there was a general decrease in J_{net} Na^+ of the Fowey fish from the start of the exposure over time, a decrease that was significantly marked in relation to the Hayle fish by 96 hours. No significant differences were detected over the course of the exposure among the Hayle fish.

The J_{net} Cl^- also showed a significant interaction between time and population (F=3.363; p=0.001; Figure 2; Table S4) and similarly, these differences were evident between the Fowey and Hayle populations at 1 hour (t=3.7; p<0.001) and 96 hours (t=2.85; p=0.005). The Hayle fish showed a relative net decrease in Cl^- at 1 hour, compared to measurements of Cl^- at 8 hours (p=0.046) and 96 hours (p<0.001). This pattern changed at 96 hours where J_{net} Cl^- significantly increased in the Hayle population in comparison to -24 hours (p =0.041), 0 hours (p =0.028), 1 hour (p<0.001) and 12 hours (p=0.004).

The Fowey fish showed no changes in plasma osmolality throughout the exposure (Table 3). On the other hand, plasma osmolality in the Hayle fish measured at the in-river time point was significantly lower in comparison to fish from the Fowey (t=3.624; p<0.001) as well as average osmolality values measured for freshwater fish (Peterson & Gilmore 1988). After depuration, osmolality in the Hayle fish was significantly higher than the Fowey population (t=2.175; p=0.031), suggesting a significant effect of the depuration period in changing the haematology of Hayle fish under clean water conditions. Plasma osmolality between exposed cohorts at 96 hours was lower in the Fowey population (296 ± 5.2 mmol/kg) in comparison to the Hayle (305 ± 4.6 mmol/kg), although this was not statistically significant (t=1.267; p=0.207). However, concentrations of Na^+ and Cl^- quantified in the blood plasma in exposed fish at 96 hours showed that Hayle fish had significantly higher concentrations of both

Na^+ (Fowey 140.9 ± 6.8 mmol/kg; Hayle 162.5 ± 6.1 mmol/kg; t=2.301; p=0.035) and Cl^- (Fowey 84.9 ± 3.6 mmol/kg; Hayle 103.4 ± 3.1 mmol/kg; t=3.854; p=0.001). Although the Hayle fish showed a slight increase in activity of Na^+/K^+-ATPase in comparison to the Fowey fish, this difference was not statistically significant (Hayle 5.59 ± 1.18 µmol; Fowey 3.27 ± 0.44 µmol, t=1.841; p=0.084).

Differences in ammonia (NH_3) excretion rates

We observed a significant difference between the Fowey and Hayle populations (F=147.9; p<0.001) and a significant interaction between time and population (F=3.98; p<0.001) for the net excretion rates of NH_3 (Figure 3). At -24 hours, no differences in net NH_3 flux were observed between the Fowey and Hayle fish. However, at 0 hours, prior to the metal exposure, Hayle fish showed a significant net increase in NH_3 excretion rates (t=2.08; p=0.039), and this became even more pronounced when the exposure was initiated (1 hour, t=4.697; p<0.001), and at every measurement of NH_3 thereafter (Table S5).

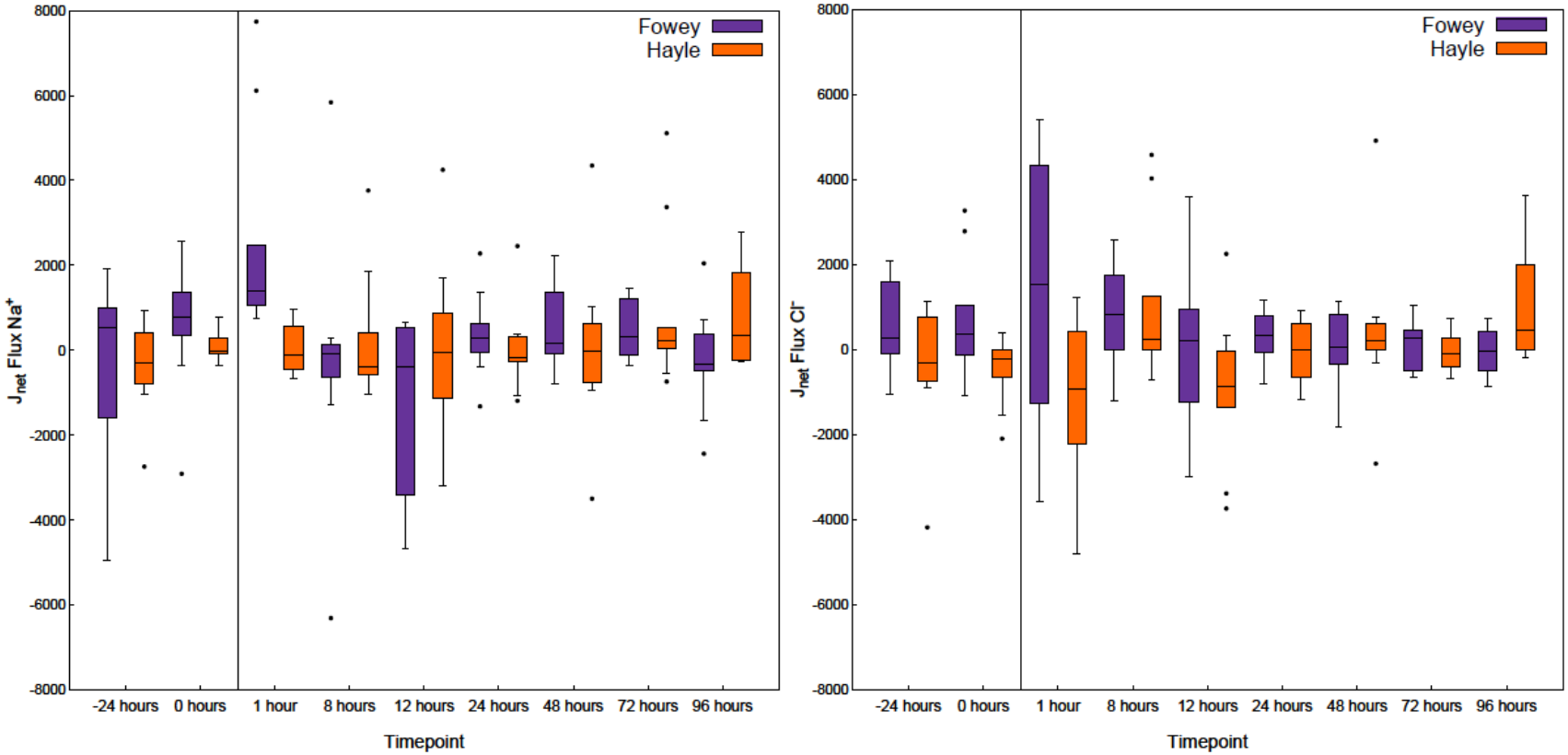


Figure 2. Plots showing the net fluxes (J_{net}) of the key ions implicated in osmoregulation: sodium (Na^+) and chloride (Cl^-). Net fluxes are represented on the y-axis as J_{net} measured as µmol/kg/h of each ion for the Fowey fish (purple) and Hayle fish (orange), and the x-axis represents the nominal time point for the measurement of each flux period. The solid vertical line represents the start of the metal exposure.

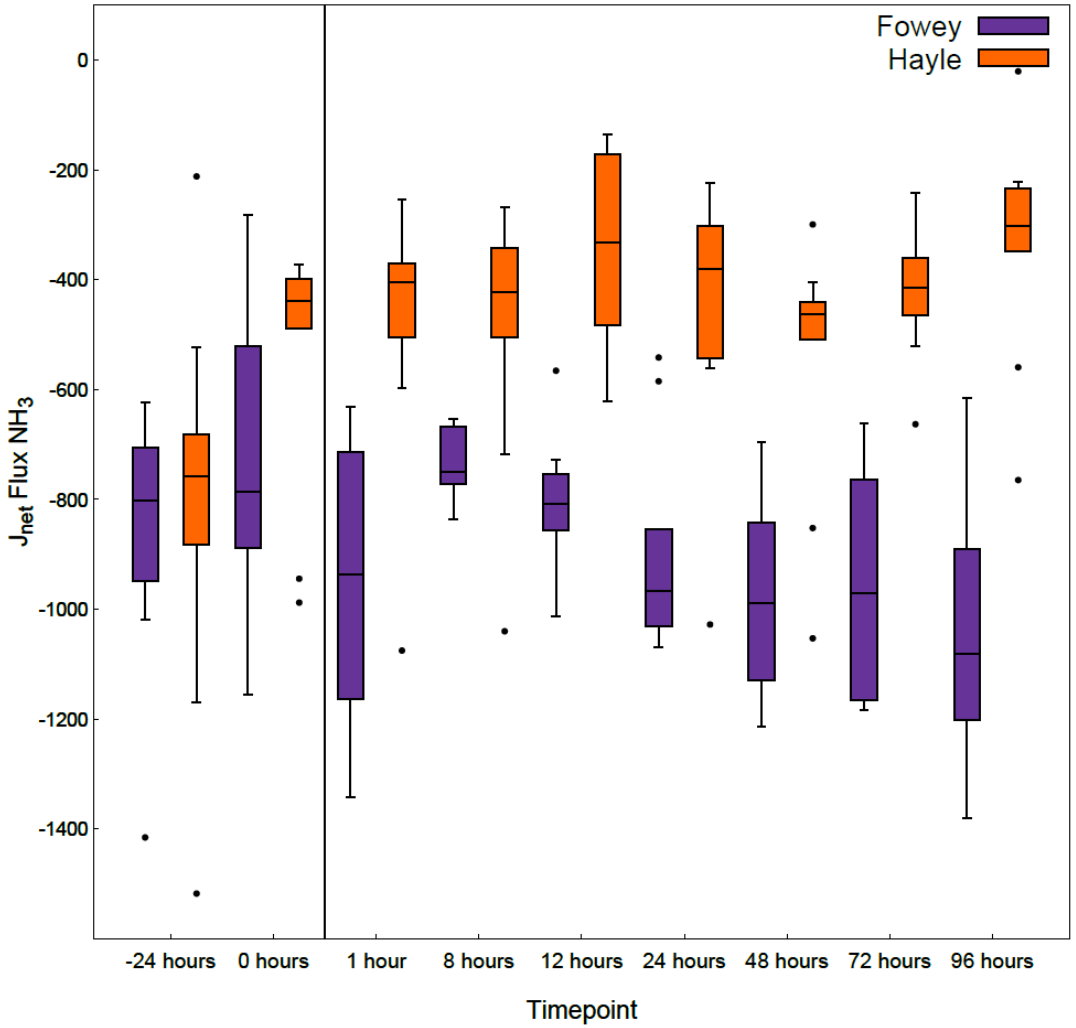


Figure 3. Plot representing changes in the net fluxes (J_{net}) of ammonia excretion (NH_3). Net fluxes are represented on the y-axis as J_{net} of NH_3 for the Fowey fish (in purple) and Hayle fish (in orange) and the x-axis represents the nominal time point for the measurement of each flux period. The solid vertical line represents the start of the metal exposure.

Metal-tissue burden analysis

1. In-river

Overall, the Hayle population exhibited patterns of high tissue metal concentrations (Table 4), suggesting that the majority of metals are being internally handled rather than being excluded through mechanisms of reduced uptake or rapidly excreted. In comparison to Fowey fish, Hayle fish showed significantly higher concentrations across a large majority of tissues and metals. This included higher concentrations of Cd in every tissue, with the largest difference observed in the kidney (Fowey 1.7 ± 0.34; Hayle 45.9 ± 4.07; $p<0.001$). With the exception of muscle, higher tissue concentrations of Cu were observed in the Hayle fish and the largest difference was observed in the liver (Fowey 26.4 ± 6.56; Hayle 801.34 ± 39.98; $p<0.001$). Except for kidney tissue, Hayle fish showed slightly higher concentrations of Zn across all tissues, but this was only statistically higher in gill (Fowey 479.76 ± 39.8; Hayle 927.97 ± 45.35; $p<0.001$) and liver (Fowey 68.85 ± 10.1; Hayle 151.90 ± 16.8; $p<0.001$). On the other hand, the Hayle population showed only slightly higher Ni concentrations; tissue differences for Ni were only significantly statistically higher in the kidney (Fowey 1.31 ± 0.38; Hayle 2.43 ± 0.31; $p=0.03$).

Table 4. *Tissue-metal burden of the Fowey and Hayle populations measured in-river*.

Metal	*Tissue*	*Fowey*	*Hayle*	*Statistical significance*
Cd	Gill	1.46 ± 0.06	10.76 ± 0.96	16.64 (p<0.001) ***
	Liver	0.44 ± 0.06	9.14 ± 1.09	5.43 (p<0.001) ***
	Kidney	1.72 ± 0.34	45.93 ± 4.07	5.71 (p<0.001) ***
	Gut	0.21 ± 0.03	2.93 ± 0.21	13.00 (p<0.001) ***
	Muscle	0.01 ± 0.00	0.07 ± 0.01	-6.827 (p<0.001) ***
Cu	Gill	2.06 ± 0.11	13.07 ± 1.27	16.56 (p<0.001) ***
	Liver	26.35 ± 6.56	801.34 ± 39.98	7.48 (p<0.001) ***
	Kidney	4.68 ± 0.52	97.92 ± 13.31	6.41 (p<0.001) ***
	Gut	3.97 ± 0.32	44.43 ± 10.53	8.65 (p<0.001) ***
	Muscle	1.18 ± 0.09	1.45 ± 0.17	-1.47 (p=0.159)
Ni	Gill	0.70 ± 0.05	1.01 ± 0.08	1.42 (p=0.158)
	Liver	0.44 ± 0.06	1.12 ± 0.09	0.48 (p=0.629)
	Kidney	1.31 ± 0.38	2.43 ± 0.31	2.29 (p=0.034) *
	Gut	0.51 ± 0.20	0.58 ± 0.11	0.27 (p=0.787)
	Muscle	0.29 ± 0.04	0.49 ± 0.14	-1.38 (p=0.183)
Zn	Gill	479.76 ± 40	927.97 ± 45.35	-7.428 (p<0.001) ***
	Liver	68.85 ± 10	151.90 ± 16.81	6.33 (p<0.001) ***
	Kidney	702.47 ± 152	524.31 ± 131.83	0.89 (p=0.387)
	Gut	335.45 ± 49	374.67 ± 29.29	0.69 (p=0.499)
	Muscle	40.04 ± 16	69.80 ± 18.02	-1.25(p= 0.228)

Results show metal accumulation of the four metals used in the exposure:Cd, Cu, Ni and Zn measured as µg/L in gill, liver, kidney, gut and muscle tissue in each population of fish. Statistical significance between the metal-burden of each population was quantified using a two-way ANOVA, accepting significance at $p<0.05$, where * ≤ 0.05; p ≤ 0.01 ** and p ≤ 0.001 ***

2. Depuration

a. Pre-exposure

With the exception of Zn in the gill, depuration did not occur in any of the Fowey fish tissues (Table 5). On the other hand, the majority of metals depurated across various tissues in the Hayle fish, with statistically significant results for: Cd in gill and liver tissue; Cu in all tissues; Ni in kidney tissue; and Zn in liver and muscle tissues. No depuration occurred for Zn in the gill and Ni in the gut of Hayle fish. As a result of the laboratory diet, concentrations of Zn significantly increased in gut tissue after depuration in both Fowey and Hayle fish.

b. During-experiment depuration in Hayle fish

(i) *Cadmium (Cd)*

In the Hayle control fish, Cd did not continue to depurate in gill and liver tissues compared to in-river concentrations (Table 6). In comparison to Fowey fish maintained in control water, Hayle fish did not show any significant depuration in kidney, gut and muscle Cd concentrations. However, kidney Cd concentrations were significantly lower at 96 hours control, compared to concentrations measured in-river ($t=3.370$; $p=0.013$), suggesting Cd concentrations in the kidney may have continued to depurate in Hayle fish.

(ii) *Copper (Cu)*

The Hayle population showed a decline in gill Cu from in-river concentrations compared to all laboratory conditions, in both control and exposed groups (24-control: $t=13.181$; $p<0.001$; 24-exposed: $t=13.4$; $p>0.001$; 96-control: $t=14.8$; $p<0.001$; 96-exposed: $t=12.5$; $p<0.001$). Compared to Fowey fish, liver Cu concentrations remained significantly higher in both Hayle control and Hayle exposed fish (Table 6), although control groups showed significantly lower concentrations at both 24 hours and 96 hours. This suggests liver Cu concentrations in the Hayle population may have continued to depurate with time. Although Cu concentrations depurated significantly in the kidney tissue of Hayle fish, significantly higher concentrations were observed when comparing the Hayle to and Fowey fish across the experiment, except at 96 hours control. Hayle fish maintained in control water had significantly lower Cu at 96 hours

compared to in-river levels, and also significantly lower Cu concentrations at 24 hours compared to 96 hours control.

(iii) *Nickel (Ni)*

Ni did not continue to depurate any further during the exposure experiment (Table 6).

(iv) *Zinc (Zn)*

Hayle fish maintained a higher concentration of gill Zn compared to Fowey fish, and this did not vary with the exposure. Both Fowey and Hayle populations showed a general decrease in muscle Zn concentrations from in-river to laboratory (Table 6).

Table 5. *Tissue-metal burden of the Fowey and Hayle populations measured in-river and a after a seven-day depuration period*.

Metal	*Tissue*	*Fowey*			*Hayle*		
		In-river	*Depuration*	*Statistical significance*	*In-river*	*Depuration*	*Statistical significance*
Cd	Gill	1.46 ± 0.06	1.40 ± 0.12	0.11 (p=1)	10.76 ± 0.96	7.22 ± 0.53	6.33 (p<0.001) ***
	Liver	0.44 ± 0.06	0.58 ± 0.09	0.09 (p=1)	9.14 ± 1.09	4.05 ± 0.62	3.27 (p=0.019) *
	Kidney	1.72 ± 0.34	1.23 ± 0.37	0.06 (p=1)	45.93 ± 4.07	29.20 ± 3.19	2.16 (p=0.259)
	Gut	0.21 ± 0.03	0.19 ± 0.03	0.63 (p=0.535)	2.93 ± 0.21	3.46 ± 0.52	-0.94 (p=0.362)
	Muscle	0.01 ± 0.00	0.01 ± 0.00	1.40 (p=0.178)	0.07 ± 0.01	0.05 ± 0.01	1.92 (p=0.0707)
Cu	Gill	2.06 ± 0.11	2.32 ± 0.07	0.39 (p=0.909)	13.07 ± 1.27	4.63 ± 0.54	12.70 (p<0.001) ***
	Liver	26.35 ± 6.56	17.35 ± 1.54	0.09 (p=1)	801.34 ± 39.98	307.85 ± 57.18	4.89 (p<0.001) ***
	Kidney	4.68 ± 0.52	4.18 ± 0.36	0.03 (p=1)	97.92 ± 13.31	44.20 ± 8.67	3.69 (p=0.005) **
	Gut	3.97 ± 0.32	4.45 ± 0.25	0.10 (p=1)	44.43 ± 10.53	10.81 ± 0.86	7.19 (p<0.001) ***
	Muscle	1.18 ± 0.09	1.18 ± 0.16	-0.04 (p=0.968)	1.45 ± 0.17	0.95 ± 0.04	2.93 (p=0.009) **
Ni	Gill	0.70 ± 0.05	1.18 ± 0.17	2.23 (p=0.244)	1.01 ± 0.08	0.80 ± 0.13	0.94 (p=0.968)
	Liver	0.44 ± 0.06	1.45 ± 0.59	1.02 (p=0.925)	1.12 ± 0.09	0.63 ± 0.06	0.82 (p=0.995)
	Kidney	1.31 ± 0.38	1.27 ± 0.41	0.08 (p=0.939)	2.43 ± 0.31	1.41 ± 0.15	2.97 (p=0.008) **
	Gut	0.51 ± 0.20	0.58 ± 0.21	-0.23 (p=0.819)	0.58 ± 0.11	0.49 ± 0.08	0.66 (p=0.517)
	Muscle	0.29 ± 0.04	0.25 ± 0.04	0.78 (p=0.445)	0.49 ± 0.14	0.23 ± 0.02	1.91 (p=0.0719)
Zn	Gill	479.76 ± 40	364.25 ± 34.76	2.19 (p=0.0423) *	927.97 ± 45.35	927.36 ± 93.20	0.01 (p=0.995)
	Liver	68.85 ± 10	55.95 ± 3.32	0.98 (p=0.938)	151.90 ± 16.81	47.83 ± 8.77	8.15 (p<0.001) ***
	Kidney	702.47 ± 152	413.96 ± 55.89	1.78 (p=0.0914)	524.31 ± 131.83	248.41 ± 31.07	2.04 (p=0.057)
	Gut	335.45 ± 49	464.51 ± 63.85	-1.61 (p=0.126)	374.67 ± 29.29	816.28 ± 74.38	-5.52 (p<0.001)
	Muscle	40.04 ± 16	12.05 ± 0.40	1.74 (p=0.0897)	69.80 ± 18.02	22.54 ± 4.56	2.54 (p=0.0204) *

Results show accumulation of the four metals used in the exposure: Cd, Cu, Ni and Zn measured as µg/L in gill, liver, kidney, gut and muscle tissue in each population of fish. Statistical significance between the metal-burden between in-river and depuration was quantified using a two-way ANOVA, accepting significance at $p<0.05$, where * ≤ 0.05; p ≤ 0.01 ** and p ≤ 0.001 ***

3. 96-hour metal mixture exposure

a. Cadmium (Cd)

Throughout the exposure, the Hayle population continued to exhibit significantly higher amounts of Cd in all tissues compared to the Fowey fish (Table 6), and the largest of these differences was observed in kidney tissue (e.g. 24 hours-exposed: $t=5.37$; $p<0.001$; 96 hours exposed: $t=6.32$; $p<0.001$). Kidney Cd concentrations increased in Hayle exposed fish compared to Hayle controls at 96 hours ($t=4.01$; $p=0.002$).

b. Copper (Cu)

With the exception of muscle tissue, Hayle fish maintained higher concentrations of Cu in all tissues in comparison to Fowey fish (Table 6), with the largest difference observed in the liver (e.g. 24 hours exposed: 7.661; $p<0.001$; 96 hours exposed: 6.008; $p<0.001$). There was a significant uptake of Cu in the liver tissue of Hayle exposed fish at both 24 hours ($t=4.97$; $p<0.001$) and 96 hours ($t=3.51$; $p=0.008$) compared to depuration; and significant uptake of Cu in the liver was also observed between Hayle exposed fish at 24 hours compared to Hayle control fish at 96 hours ($t=3.47$; $p=0.008$). There was a significant uptake of Cu in the kidney in Hayle exposed fish at 96 hours compared to the Hayle control group ($t=3.06$; $p=0.033$). Both Hayle control and Hayle exposed fish showed significantly lower concentrations of gut Cu compared to in-river concentrations (24 control: $t=6.48$; $p<0.001$; 24 exposed: $t=6.29$; $p<0.001$; 96 control: $t=7.63$; $p<0.001$; 96 exposed: $t=6.19$; $p=<0.001$). Fowey exposed fish showed an increase in gill Cu concentrations at 96 hours, compared to all Fowey control groups (24 control: $t=3.58$; $p=0.007$; 96 control: $t=2.976$; $p=0.042$).

c. Nickel (Ni)

The Fowey fish showed significant alterations in gill Ni concentrations associated with the exposure (Table 6): at 96 hours, exposed fish accumulated significantly higher levels of Ni in gill tissue, compared to control groups (24 control: $t=3.19$; $p=0.026$; 96 control: $t=3.137$; $p=0.028$) and compared to uptake at 24 hours exposed ($t=2.931$; $p=0.048$). Similarly, there were significant increases in concentrations of Ni in the livers of Fowey exposed fish at 96 hours compared to Fowey measurements in-river ($t=3.73$; $p=0.005$), and also a

significant increase from 24 hours exposed (t=3.297; p=0.018). Ni concentrations in the kidney remained significantly higher in the Hayle population in comparison to the Fowey population, with no significant alterations between control and exposed cohorts (24 hours: t=0.35; p=0.732; 96 hours: t=1.548; p=0.139). In comparison to depuration, both Hayle and Fowey fish showed a significant uptake of Ni in muscle tissue at 96 hours in both control (t=3.03; p=0.045) and exposed (t=3.03; p=0.042) cohorts.

d. Zinc (Zn)

Hayle fish maintained higher concentrations of gill Zn compared to Fowey fish (t=11.92; p<0.001), and this did not vary with the exposure (Table 6). At 96 hours, Fowey exposed fish showed significantly higher gill Zn compared to Fowey controls (t=4.29; p<0.001). Hayle exposed fish accumulated significantly more kidney Zn compared to Hayle controls (t=2.227; p=0.04). At 96 hours, Fowey exposed fish showed higher concentrations of Zn in liver tissue compared to Hayle exposed (t=4.419; p<0.001); this increase was statistically higher when compared to all other control measurements in the Fowey population (24 control: t=4.65; p<0.001; 96 control: t=4.29; p<0.001). In comparison, Hayle fish showed significantly decreased concentrations of liver Zn from in-river, when compared to all other time points, regardless of control or exposed groups (24 control: t=6.41; p<0.001; 24 exposed: t=5.3; p<0.001; 96-control: t=7.77; p<0.001; 96-exposed: t=7.06; p<0.001). Both Fowey and Hayle populations showed an increase in gut Zn concentrations from in-river compared to laboratory conditions (F=9.63; p<0.001), and uptake of gut Zn was significantly higher in the Hayle population (F=24.11; p<0.001).

Table 6. *Tissue-metal burden of the Fowey and Hayle populations measured during the exposure at 24 hours and 96 hours*.

Metal	*Tissue*	*Fowey*				*Hayle*			
		24 hours control	*24 hours exposed*	*96 hours control*	*96 hours exposed*	*24 hours control*	*24 hours exposed*	*96 hours control*	*96 hours exposed*
Cd	Gill	1.22 ± 0.09	1.63 ± 0.09	1.25 ± 0.12	1.43 ± 0.14	6.57 ± 0.57	6.96 ± 0.36	6.06 ± 0.16	5.80 ± 0.36
	Liver	0.73 ± 0.12	0.86 ± 0.12	0.83 ± 0.09	1.44 ± 0.30	7.85 ± 1.96	11.29 ± 2.24	6.54 ± 0.92	9.35 ± 1.74
	Kidney	1.60 ± 0.26	1.40 ± 0.36	1.74 ± 0.34	1.98 ± 0.47	52.62 ± 14.98	42.94 ± 6.30	19.85 ± 3.91	50.87 ± 7.23
	Gut	0.25 ± 0.02	0.24 ± 0.03	0.34 ± 0.04	0.28 ± 0.02	3.12 ± 0.54	4.33 ± 0.56	2.80 ± 0.29	3.35 ± 0.31
	Muscle	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.06 ± 0.01	0.08 ± 0.01	0.05 ± 0.00	0.06 ± 0.01
Cu	Gill	2.46 ± 0.15	3.77 ± 0.20	2.32 ± 0.14	4.84 ± 0.62	4.31 ± 0.26	4.16 ± 0.21	3.23 ± 0.17	4.75 ± 0.36
	Liver	25.95 ± 3.67	36.10 ± 10.75	32.27 ± 2.47	55.66 ± 12.14	718.76 ± 161.64	808.61 ± 118.67	459.20 ± 80.35	661.53 ± 92.68
	Kidney	5.74 ± 0.60	4.54 ± 0.49	4.92 ± 0.52	6.16 ± 0.48	80.91 ± 25.57	62.82 ± 15.99	27.06 ± 6.55	71.65 ± 8.06
	Gut	4.27 ± 0.13	3.99 ± 0.12	6.58 ± 0.45	5.47 ± 0.21	14.14 ± 2.44	15.01 ± 2.84	8.73 ± 0.59	15.45 ± 2.18
	Muscle	1.08 ± 0.05	1.08 ± 0.07	1.12 ± 0.17	1.37 ± 0.09	1.26 ± 0.09	1.26 ± 0.11	1.00 ± 0.06	1.17 ± 0.12
Ni	Gill	1.09 ± 0.09	1.10 ± 0.14	1.05 ± 0.14	1.73 ± 0.39	1.08 ± 0.16	0.79 ± 0.04	0.70 ± 0.03	0.73 ± 0.05
	Liver	1.81 ± 0.59	1.14 ± 0.26	1.52 ± 0.35	3.10 ± 0.96	1.11 ± 0.13	1.93 ± 0.45	1.04 ± 0.15	1.11 ± 0.15
	Kidney	1.15 ± 0.31	1.95 ± 0.96	1.12 ± 0.23	1.78 ± 0.77	3.14 ± 0.63	3.56 ± 1.03	1.36 ± 0.33	1.88 ± 0.28
	Gut	0.44 ± 0.11	0.49 ± 0.14	0.47 ± 0.06	0.47 ± 0.07	0.43 ± 0.08	0.53 ± 0.16	0.50 ± 0.11	0.34 ± 0.05
	Muscle	0.45 ± 0.05	0.32 ± 0.04	0.57 ± 0.10	0.50 ± 0.07	0.40 ± 0.07	0.40 ± 0.06	0.33 ± 0.04	0.41 ± 0.10
Zn	Gill	460.16 ± 51.24	476.00 ± 43.43	439.92 ± 37.82	454.12 ± 33.04	850.10 ± 112.79	1087.93 ± 69.90	975.46 ± 142.32	939.19 ± 79.15
	Liver	58.72 ± 2.23	81.32 ± 7.04	63.28 ± 6.50	118.15 ± 14.48	70.06 ± 8.15	84.16 ± 9.38	52.58 ± 4.41	61.69 ± 6.13
	Kidney	507.49 ± 52.30	534.00 ± 128.75	662.63 ± 116.34	657.77 ± 152.14	401.68 ± 86.08	531.21 ± 111.21	194.22 ± 45.73	453.89 ± 104.44
	Gut	535.58 ± 59.94	554.06 ± 47.08	813.73 ± 110.94	699.52 ± 83.42	821.51 ± 95.24	931.84 ± 84.73	913.28 ± 160.63	966.05 ± 61.22
	Muscle	16.54 ± 3.00	13.65 ± 1.54	48.85 ± 12.04	48.65 ± 22.15	41.93 ± 15.20	33.85 ± 5.33	20.64 ± 2.70	49.15 ± 19.69

Results show metal accumulation of the four metals used in the exposure: Cd, Cu, Ni andZn measured as µg/L in gill, liver, kidney, gut and muscle tissue in each population of fish. Statistical significance between the metal-burden between in-river and depuration was quantified using a two-way ANOVA, accepting significance at $p<0.05$, where * ≤ 0.05; p ≤ 0.01 ** and p ≤ 0.001 ***

DISCUSSION

The mechanisms underlying the metal-handling physiology of fish determine their ability to regulate the uptake, storage and detoxification of metals. In this study, we investigated the metal accumulation, osmoregulation and ammonia excretion processes of two populations of trout, one from a chronically contaminated river (River Hayle) and another from a reference control river (River Fowey). Amongst the various physiological parameters that were measured, differences in several condition factors, osmoregulation, ammonia excretion and metal-tissue accumulation patterns were observed.

Several of the physiological condition factors suggest that Hayle fish are not experiencing negative toxic effects as a result of their metal-rich environment. In comparison to the Fowey population, low HSI values in Hayle trout were consistently observed. As a proxy for glycogen reserves, smaller liver sizes (low HSIs) are frequently cited as a sign of poor diet or starvation in fish (Shoemaker *et al.* 2003; Rios *et al.* 2006). However, because the liver is also an important organ involved in metal storage and detoxification, high values for the HSI parameter have also be used as an indicator of hepatic damage caused by chronic metal exposure. For example, increases in HSI correlate with metal-induced liver hypertrophy and hyperplasia (Khan 2003; Facey *et al.* 2005; van Dyk *et al.* 2012; Dragun *et al.* 2013). The HSI values observed here are indicative that Hayle fish are not experiencing hepatic damage, and also point to mechanisms of efficient metal handling within the liver (see discussion on metal-tissue burden data).

HCT measured between the two fish populations suggested that Hayle fish were relatively unaffected haematologically, both in respect to the metal concentrations experienced in-river, and also when challenged by acute metal exposure in the laboratory, whereas the Fowey fish showed distinct adjustments related to the metal exposure. Fish blood is sensitive to pollution-induced stress, and measurements of blood parameters such as HCT are valuable in monitoring the oxygen carrying capacity of the blood, as well as indicating a general stress response to pollutants (Larsson *et al.* 1985; Gill & Epple 1993; Al-Asgah *et al.* 2015). Significant increases in HCT are frequently observed as a haematological response in fish exposed to a wide range of metals (Majewski & Giles 1981; Wilson & Taylor 1993; Safahieh *et al.* 2010). Such increases in HCT

have been associated with the damage to, and swelling of, erythrocytes caused by metals (see review by Witeska 2013). Increased HCT has also been correlated with an increased production of red blood cells in response to hypoxia-induced stress (Holeton & Randall 1967) and lowered oxygen tension (Soivio *et al.* 1974), both of which are indirect effects of metal exposure (Skidmore 1970; Lappivaara *et al.* 1995; Heerden *et al.* 2004).

Observations of the adjustments in osmoregulatory ions also indicated differences between the metal-handling capacity of Hayle and Fowey fish. Unexpectedly, at 1 hour into the exposure, the Fowey fish showed a net increase of Na^+ and Cl^- relative to the Hayle fish. As metal toxicity causes an inhibition of osmoregulatory processes, it was expected that net decreases in the fluxes of these ions would be observed in metal-naïve fish (McDonald *et al.* 1989; McDonald & Wood 1993; Wood 1992), as has also been shown in other studies (Suvetha *et al.* 2010; Clemow & Wilkie 2015; Chowdhury *et al.* 2016). This result can be explained by overcompensation in response to an initial inhibition of Na^+ and Cl^- uptake in Fowey fish on first exposure to the metals (Laughlin *et al.* 1981; Calabrese & Baldwin 1997). After this over-compensation, ionoregulatory processes were negatively affected as would be expected in metal-naïve fish, such that a significant decrease in the net fluxes of these ions was observed in the Fowey fish by 96 hours. Overcompensatory responses in ionoregulation have seldom been reported in the literature, but Na^+/K^+-ATPase activity in the gills of Coho salmon (*Oncorhynchus kisutch*) exposed to xenobiotics (Tetrachloroguaicol; TeCG), showed increased activity, when it was expected that the treatment would cause an inhibition (Yang & Randall 1996). Similarly, in mud crabs (*Scylla serrata*) challenged by hypo- and hyper-salinity changes, the transport rate of K^+, Na^+ and Ca^+ ions across the cell membrane showed rapid over-compensations in both treatments (Roy & Bhoite 2016). Such over-compensations represent a significant area for further investigation on the effects of toxicants on osmoregulation.

Plasma osmolality measured in-river was significantly lower in the Hayle fish compared to fish from the Fowey. Low plasma osmolality is indicative of an overall effect of metals on the functioning of ionoregulation, and is measured as a biomarker of metal toxicity in fish (Monteiro *et al.* 2005; Reynders *et al.* 2006). However, in Cd-tolerant Nile tilapia (*Oreochromis niloticus*) exposed to

extremely high concentrations of Cd, no changes in plasma Na^+ were evidenced (Garcia-Santos *et al.* 2006), and in Cu-acclimated fathead minnows (*Pimephales promelas*), whole-body Na^+ concentrations remained unchanged, although short-lived and moderate reductions in Na^+ were observed on re-exposure to Cu (Tate-Boldt & Kolok 2008). In addition, other proteins and ions could be responsible for the differences in plasma osmolality observed here. For example, increases in bicarbonate (HCO_3-) as a result of acid-base balance disturbances (Spry & Wood 1985; Wood *et al.* 1988; Wang *et al.* 1998) may also be implicated in the low values of plasma osmolality observed in the Fowey fish at 96 hours (Nonnotte & Truchot 1990; Lefevre *et al.* 2012). Although not measured, differences in the acid-base regulation of HCO_3- in response to metal stress have been observed in fish (Larsen *et al.* 1997; Wood *et al.* 1988). Furthermore, Cu and Al are both known to inhibit intracellular carbonic anhydrase in the ionocytes; this enzyme hydrates CO^2 so as to produce the H^+ ions that are normally exchanged against Na^+ and HCO_3^- (Staurnes *et al.* 1984; Wang *et al.* 1998; Wood *et al.* 2012). Acid-base regulation may be another physiological mechanism that is highly regulated in Hayle fish.

Taken together, these results suggest that Hayle fish do not experience metal-induced stress associated with osmoregulation, implying that gill-specific mechanisms likely play a role in regulating the negative effects of chronic metal exposure. Although not significant, the increased activity of Na^+/K^+-ATPase suggests that this could represent a significant mechanism in maintaining ionoregulatory balance. Alternatively, other proteins not measured here, such as V-type H^+ATPase (Grosell & Wood 2002; Goss *et al.* 2011) and regulation of metal-specific carriers (e.g. Ctr1, DMT1 and Cu-ATPase for Cu; ZIP and ZnT for Zn; Bury *et al.* 2003; Minghetti *et al.* 2008; Mingehtti *et al.* 2010; Hogstrand 2012) could also account for the differences observed. Identifying the activity of these proteins in the gill tissue of Hayle fish represents a significant avenue for further research.

Distinctive differences in ammonia (NH_3) excretion rates between the two populations were observed, and interestingly, these differences were apparent before initiation of the metal exposure. NH_3 is a natural waste product of nitrogen metabolism and is excreted through the gill via proton exchange and rhesus proteins located at the gill surface (Walsh *et al.* 2001 Nawata *et al.* 2007;

Wright & Wood 2009). Metals inhibit NH_3 excretion by blocking apical H^+ extrusion, interfering with H^+-ATPase and through disruption of carbonic anhydrase in the gill (Vitale *et al.* 1999; Ceyhun *et al.* 2011; Zimmer *et al.* 2012; Ito *et al.* 2013; Wood *et al.* 2014). Therefore, sub-lethal exposures to metals typically show an increase in the net rate production of NH_3 (Wilson & Taylor 1993; Beaumont *et al.* 2003; Blanchard and Grosell, 2006). Excessive accumulation of NH_3 is extremely toxic to fish (Randall & Tsui 2002), and therefore it is possible that Hayle fish have developed efficient NH_3 excretion processes in order to survive in their metal-rich environment.

This could also suggest that the Hayle fish may have constitutively up-regulated mechanisms in controlling the efficiency of the NH_3 excretion process. The transport of NH_3 across the gill is facilitated by Rhesus (Rh) glycoproteins, which actively transport NH_3 across the gill epithelium (Nakada *et al.* 2007; Hung *et al.* 2008). It has been suggested that metals may disrupt the activity of Rh proteins (Lim *et al.* 2015), but compensatory increases in Rh glycoprotein activity have also been implicated in responses to toxicant and ammonia stress (Wood *et al.* 2014; Wright *et al.* 2014; Sinha *et al.* 2016). Although not measured here, increased activity of Rh glycoproteins represent a potential mechanism of efficient NH_3 excretion in Hayle fish in order to counteract the toxicity of metals in their native environment.

An alternative hypothesis for the increased NH_3 excretion rates observed in Hayle fish is differential regulation of gluconeogenesis. The liver is an important site of gluconeogenesis in fish, with amino acids representing an important source of carbon (Ballantyne 2001). Because the fish in this experiment had fasted for 24 hours before measurements of NH_3 fluxes, alternative methods of glycogenesis could indicate different rates of protein metabolism between the two populations. In experiments assessing the physiology of rainbow trout (*O. mykiss*), NH_3 excretion rates showed a four-fold elevation in the use of protein as an aerobic fuel for routine metabolism in satiation-fed individuals (50-70%), compared to fasted fish (Aslop & Wood 1997). Shifts in NH_3 excretion have also been associated with increased metabolic demands associated with swimming activity, highlighting the reliance on protein reserves as an aerobic fuel in fish (Aslop *et al.* 1999). For example, protein use represented 30% of instantaneous aerobic fuel in sustained swimming activity in rainbow trout (*O. mykiss*) (Lauff &

Wood 1996). Associated with the low values of HSI observed in the Hayle fish (potentially suggesting reduced glycogen reserves), protein metabolism may represent a significant fuel for maintaining metabolic activity in these fish. However, because we did not quantify oxygen consumption, differences in metabolic rates between the two fish populations cannot be accurately assessed here, but should be explored in future research.

Differential patterns of metal uptake, accumulation and storage between Fowey and Hayle fish

In terms of characterising the most probable metal-handling routes and the mechanisms of metal detoxification and excretion of Hayle and Fowey fish, general patterns were observed for each metal. In particular, there appeared to be marked differences between accumulation of non-essential (Cd) and essential (Cu, Ni, Zn) metals between the two populations. Significantly, metal-handling routes in Hayle fish appear to be highly modified in comparison to fish from the River Fowey.

In-river Cd was significantly higher in all tissues in the Hayle fish, suggesting the potential importance of a metal-detoxified fraction (MDF) in negating the toxic impacts of this metal. Overall, acclimation to chronic Cd exposure in fish is known to involve a variety of defense and detoxification processes, including antioxidants (Basha & Rani 2003), metallothioneins (De Smet *et al.* 2001; Lange *et al.* 2002), glutathione (Rana & Singh 1996), calcium granules (Campbell *et al.* 2005) and heat shock proteins (HSPs; Wang & Rainbow 2013). Exposures of juvenile rainbow trout (*Oncorhynchus mykiss*) showed that gill Cd concentrations were 20- to 40-fold greater than levels expected to cause mortality during acute exposure as predicted by the gill-binding model (Hollis *et al.* 1999). This points to the significance of the MDF in Cd toxicity and handling in fish.

After entering the blood stream, Cd is rapidly absorbed from the plasma by various internal organs, but the highest percentage absorption is known to occur in liver and kidney (McGeer *et al.* 2012). After the first week of depuration, although a portion of Cd depurated from the gill and liver tissues of Hayle fish, concentrations remained significantly higher in the kidney in comparison to Fowey fish. Kidney tissue frequently accumulates the highest concentrations of

Cd (De Smet *et al.* 2001; Andres *et al.* 2000), and HSPs are purported to be the main sub-cellular binding proteins involved (Kraemer *et al.* 2005; Campbell *et al.* 2008; Kamunde 2009). Hayle fish maintained in control water conditions showed a significant decrease in kidney Cd concentrations, compared to in-river levels, suggesting that a proportion of Cd was being moved from a metal-sensitive fraction (MSF) to a metal-detoxified fraction (MDF). Cd is an extremely potent inducer of metallothioneins and these proteins likely represent a significant constituent of the MDF in Hayle fish, particularly in the kidney (Klaassen *et al.* 1999; Dudley *et al.* 1985; Van Campenhout *et al.* 2004).

Across all tissues, Cd concentrations in Hayle fish did not depurate to the levels observed in the Fowey fish, suggesting that Cd remains bound to the MDF as a non-toxic form throughout the lifetimes of Hayle fish. Interestingly, in the previous study by Uren-Webster *et al.* (2013), which measured metal-tissue concentrations in adult Hayle trout (weight 142g ± 21; fork-length 21cm; ± 36), extremely large differences were observed in the metal-accumulation of Cd in comparison to the younger (0+ fry) trout that were used in this study: Gill: 154 ng/mg, 10.1 ng/mg (15-fold change); Kidney 581 ng/mg, 45 ng/mg (13-fold change); Liver: 17 ng/mg, 1.86 ng/mg (1.8 fold-change); Gut: 16 ng/mg, 2.9 ng/mg (5.4 fold-change). This is highly indicative of long-term metal-storage processes for Cd in Hayle fish over time, particularly in the gill and kidney tissues. However, it is not possible to determine whether the efficiency of the Hayle fish's capacity to bind Cd is a stable strategy for longevity. For example, chronic sub-lethal exposure to Cd over time can lead to degradation in tissue ultrastructure of gills, liver, and kidney (Randi *et al.* 1996; Thophon *et al.* 2003).

Several studies have demonstrated that Cu-acclimated fish demonstrate enhanced Cu tolerance on re-exposure (Grosell *et al.* 1997; Laurén & McDonald 1987; Dang *et al.* 2009); for example, metal-naïve rainbow trout experienced an 80% mortality rate when exposed to 91 µg/L of Cu over 96-hours, whereas trout acclimated to 36 µg/L Cu for 56 days prior to the acute exposure showed only a 20% mortality rate (Hansen *et al.* 2002). Here, we observed tissue-specific differences between the handling of Cu between the Fowey and Hayle fish, and a high efficiency of Cu detoxification in the Hayle fish.

The gill plays a key role in Cu homeostasis in fish (Kamunde *et al.* 2002b), whereas the liver is the primary organ involved in Cu metabolism, accounting for 25–60% of whole-body Cu (Grosell 2012). Cu concentrations showed significant patterns of accumulation in Fowey gill tissue, suggesting that these fish are experiencing acute Cu exposure at this site, and are potentially negating the effects of Cu by gill homeostatic mechanisms. Gill metallothionein is regulated in response to Cu exposure (Dang *et al.* 1999; De Boeck *et al.* 2003), and hypoxia-induced genes are known to play a role in regulating the toxic effects at high concentrations (Heerden *et al.* 2004). On the other hand, the Hayle fish showed no such accumulation at the gill; instead showing significant increases of Cu concentrations in the liver and kidney tissues. In pre-exposed Cu-acclimated juvenile rainbow trout (*O. mykiss*), re-exposure to Cu showed decreased patterns of Cu accumulation in gill tissue (Kamunde *et al.* 2002b).

Cu showed the highest accumulation in liver tissue of Hayle fish in comparison to the Fowey fish, which showed no such accumulation patterns. High concentrations of hepatic Cu are due to efficient clearance of plasma Cu (Grosell *et al.* 1997; Grosell *et al.* 2001). The heat-stable cytosolic peptides and proteins fraction, such as metallothioneins, have been shown to play a major role in hepatic metabolism of Cu (Bremner 1987; Giguère, *et al.* 2006; Lange *et al.* 2002). Interestingly, several studies point to a relatively minor role of metallothioneins as Cu storage proteins in fish (Hogstrand *et al*. 1991; Grosell *et al.* 1997, 1998). Once bound in the liver, excretion of both waterborne and dietary Cu is known to occur via heptobiliary routes (Grosell 2012), most likely involving proteins such as *atp7b* (Minghetti *et al.* 2010; da Silva *et al.* 2014). Binding of Cu to metallothioneins in the kidney also contributes a role in Cu homeostasis (Hollis *et al.* 2001; De Boeck *et al.* 2003) and excretion is known to occur through renal excretion of urine (Grosell *et al.* 1998). This correlates with the Cu accumulation patterns observed in the liver and kidney tissues of exposed Hayle fish, and the significant decrease of kidney and liver Cu concentrations in Hayle control cohorts over time. The implication is that Cu is being transported to both the liver and kidney for detoxification and excretion. Indeed, Uren-Webster *et al.* (2013) showed upregulated genes involved in glutathione synthesis in the gill, which could account for the efficient movement of Cu by these proteins from gill to liver and kidney tissues.

The acute toxicity of Ni is relatively low in comparison to most other metals, its toxicity to fish typically ranked lower than Cd, Cu, and likely equally as toxic as Zn (Pyle & Couture 2012). The patterns of Ni accumulation and depuration suggest that Ni had no effect on Hayle fish gill accumulation. On the other hand, Fowey fish showed a significant uptake of Ni in the gill. Ni typically shows little to no effect on branchial ionoregulatory disturbances; in an exposure of 15.3 mg/l of Ni to rainbow trout, no disruptions in the net fluxes of branchial ions were shown (Pane *et al.* 2003), whereas an exposure to 9.7–10.7 mg Ni L^{-1} showed acute respiratory disturbances via ultrastructural damage to the respiratory epithelium of the gill (Pane *et al.* 2004). This suggests that, as shown with Cu handling, Fowey fish are regulating mechanisms of Ni toxicity at the gill site but Ni is unlikely to be contributing to the osmoregulatory adjustments that were observed.

Hayle fish showed distinct patterns of Ni accumulation in the kidney. Fish maintained in control cohorts showed significant depuration over time, suggesting a role of the kidney in both Ni homeostasis and excretion. Although comparatively less is known about Ni handling in fish, several studies have demonstrated the preferential accumulation of Ni in the kidney (Ptashynski *et al.* 2001; Pane *et al.* 2004; Chowdhury *et al*. 2008; Palermo *et al.* 2015), although little is known about the exact mechanisms of protein handling. Furthermore, studies of chronic Ni exposure in rainbow trout (*O. mykiss*) showed that renal excretion is a possible method of Ni secretion (Pane *et al.* 2005, Pane *et al.* 2006ab) and, interestingly, chronically exposed fish showed a reduced reabsorption rate in comparison to metal-naïve fish (Pane *et al.* 2006b). In rainbow trout pre-exposed to 442 µg Ni L^{-1} for 36 days, acclimated fish were shown to be well-protected against Ni-induced Mg^{2+} antagonism, whereas naïve trout showed extreme biological disturbances in kidney physiology (Pane *et al.* 2005). The kidney therefore likely plays a vital role in reducing the impacts of Ni toxicity in Hayle fish, although the precise mechanisms of Ni handling and homeostasis cannot be ascertained.

In comparison to the tissue accumulation patterns of Cu and Ni, Zn showed accumulation in both the gills and the liver of the Fowey fish. This suggests that Zn requirements in the Fowey fish are different from those of the other two essential metals measured here. Zn is the second most abundant trace metal in

vertebrates, with essential functions in a wide variety of processes including the metabolism of nucleic acids, carbohydrates and lipids, and vital roles in immune system function, neurotransmission and cell signaling (Hogstrand *et al* 1991; Hogstrand 2012). This high requirement for Zn could account for the different patterns of tissue accumulation observed in the Fowey fish.

At low concentrations, Zn is actively taken up from water via the gill (Hogstrand & Wood 1996; Santore *et al.* 2002), in order to negate adverse health effects caused by Zn deficiency. Thus, detrimental health effects of Zn deficiency can arise if Zn is not handled effectively at the primary route of uptake (Bury *et al.* 2003). In an experiment of rainbow trout (*Oncorhynchus mykiss*) fed on either excessive or deficient Zn diets, the waterborne contribution of Zn was shown to be important; as high at 57% even when dietary Zn was adequate, and 100% when dietary Zn was deficient (Spry *et al.* 1988). Furthermore, the highest fold change in whole-body Zn content occurred in the Zn-deficient cohort (Spry *et al.* 1988). Indeed, uptake of Zn in the liver of exposed Fowey fish was significantly higher at 96 hours compared to the Hayle fish. We therefore suggest that Fowey fish may be actively accumulating Zn in an effort to avoid deficiency, rather than having to regulate the toxicity of this essential metal. Zn in deficiency may therefore represent a health concern in this population of fish.

On the other hand, Zn appeared to be saturated in the gills of Hayle fish, as no depuration occurred throughout the experiment. However, gill Zn concentrations measured in this study were only 1.2-fold higher than concentrations measured in adult Hayle trout (Uren-Webster *et al.* 2013) (Gill Zn: 1132 ng/mg; 927 ng/mg). This suggests modified mechanisms of Zn turnover at the gill site and/or efficient mechanisms of Zn excretion. In freshwater fish, Zn uptake through the gill can contribute to 50% of total Zn absorption if Zn concentrations are high (Hogstrand 2012). Once within the gill, handling of Zn occurs primarily through two main families of Zn transporters: Slc30 (Zn transporters, ZNTs) and Slc39 (Zrt IRT-related proteins, ZIPs) (Qui *et al.* 2007; Zheng *et al.* 2008), whilst binding occurs in association with glutathione and metallothioneins (Lange *et al.* 2002; Van Campenhout *et al.* 2003; Hansen *et al.* 2007). Quantifying the gene expression and activity of these proteins in Hayle fish would shed light on the ability of Hayle fish to regulate Zn toxicity in the gills.

Hayle fish also showed significant accumulation of Zn in the liver, as has been observed in hepatic tissue in other fish species (Holcombe *et al.* 1979; Hogstrand *et al.* 1991; Köck & Bucher 1997). Indeed, the liver plays an important role in Zn detoxification and storage (Hogstrand 2012); for example, in carp (*Cyprinus carpio*), over 30% of cytosolic Zn was bound to metallothioneins in the liver, whilst this accounted for only 2% in the kidneys (Van Campenhout *et al.* 2004). However, it is thought that hepatic excretion may only account for a small proportion of Zn elimination; for example, in rainbow trout (*Oncorhynchus mykiss*), less than 1% of injected or dietary Zn was found in the bile (Chowdhury *et al.* 2003), leading to suggestions that Zn may be excreted via the gut (Hogstrand 2012). However, in this study we saw a significant depuration of Zn in the liver of Hayle fish and no associated patterns of Zn accumulation in gut tissue.

In conclusion, Hayle fish appear to have unique and synergistic mechanisms in place to cope with chronic and acute metal toxicity in terms of osmoregulatory processes, ammonia excretion methods and in the binding and detoxifying of essential and non-essential metals. Over generations of exposure to metals, Hayle fish have likely adapted by compensating for losses of ions by increasing a number of key transporters involved in maintaining osmoregulatory homeostasis. Adjusted mechanisms of internal metal handling exist in these metal-tolerant fish, both through binding of metals to the metal-detoxified fraction (MDF) and also through binding of metals to the metal-sensitive fraction (MSF) followed by efficient excretion of these metals. Together, these physiological processes point to adaptive mechanisms involved in the metal-tolerance of Hayle fish, allowing them to live in a chronically metal polluted environment.

ACKNOWLEDGEMENTS

We would like to thank the University of Exeter, Environment Agency and Westcountry Rivers Trust for funding this work. Thanks to Matt Healey, Giles Ricard and Dr Bruce Stockley from the Westcountry Rivers Trust and also to Jan and Phil Shears for organising the electrofishing and assisting with fieldwork for the fish collection. Thanks to everyone in Lab 201 and the Aquatic Resource Centre for assisting with the experimental exposure, especially Dr Gregory Paull, Steve Cooper, and John Dowdle, and also to Dr Anke Lange, Dr

Jennifer Fitzgerald and Lauren Laing for assistance with sampling and experimental design. A special thank you to Dr Jane Usher for laboratory and experimental assistance throughout the experiment. Thanks also to Dr Andy Cakebread, University King's College London for the ICP-MS analysis and Dr Rob Ellis for assistance with analysis on the Dionex.

REFERENCES

Al-Asgah, N. A., Abdel-Warith, A.-W. A., Younis, E.-S. M., & Allam, H. Y. (2015). Haematological and biochemical parameters and tissue accumulations of cadmium in *Oreochromis niloticus* exposed to various concentrations of cadmium chloride. *Saudi Journal of Biological Sciences*, **22**(5), 543–50.

Alsop, D. H., Kieffer, J. D., & Wood, C. M. (1999). The effects of temperature and swimming speed on instantaneous fuel use and nitrogenous waste excretion of the Nile tilapia. *Physiological and Biochemical Zoology : PBZ*, **72**(4), 474–83.

Alsop, D., & Wood, C. (1997). The interactive effects of feeding and exercise on oxygen consumption, swimming performance and protein usage in juvenile rainbow trout (*Oncorhynchus mykiss*). *Journal of Experimental Biology*, **200**(17), 2337-2346.

Andres, S., Ribeyre, F., Tourencq, J.-N., & Boudou, A. (2000). Interspecific comparison of cadmium and zinc contamination in the organs of four fish species along a polymetallic pollution gradient (Lot River, France). *Science of the Total Environment*, **248**(1), 11–25.

Atli, G., & Canli, M. (2007). Enzymatic responses to metal exposures in a freshwater fish *Oreochromis niloticus*. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, **145**(2), 282–287.

Atli, G., & Canli, M. (2013). Metals (Ag^{+}, Cd^{2+}, Cr^{6+}) affect ATPase activity in the gill, kidney, and muscle of freshwater fish *Oreochromis niloticus* following acute and chronic exposures. *Environmental Toxicology*, **28**(12), 707–717.

Baatrup, E. (1991). Structural and functional effects of heavy metals on the nervous system, including sense organs, of fish. *Comparative Biochemistry and Physiology Part C: Comparative Pharmacology*, **100**(1–2), 253–257.

Ballantyne, J.S. Amino acid metabolism. 2001. In Wright, P.A. & Anderson, P.M. (eds.). *Nitrogen Excretion*. California, London: Academic Press. pp. 77-99.

Basha, P. S., & Rani, A. U. (2003). Cadmium-induced antioxidant defence mechanism in freshwater teleost *Oreochromis mossambicus* (Tilapia). *Ecotoxicology and Environmental Safety*, **56**(2), 218–221.

Beaumont, M. W., Butler, P. J., & Taylor, E. W. (2003). Exposure of brown trout *Salmo trutta* to a sublethal concentration of copper in soft acidic water: effects upon gas exchange and ammonia accumulation. *The Journal of Experimental Biology*, **206**(1), 153–62.

Benoit, D. A., Leonard, E. N., Christensen, G. M., & Fiandt, J. T. (1976). Toxic Effects of Cadmium on Three Generations of Brook Trout (*Salvelinus fontinalis*). *Transactions of the American Fisheries Society*, **105**(4), 550–560.

Blanchard, J., & Grosell, M. (2006). Copper toxicity across salinities from freshwater to seawater in the euryhaline fish *Fundulus heteroclitus*: Is copper an ionoregulatory toxicant in high salinities? *Aquatic Toxicology*, **80**(2), 131–139.

Bougas, B., Normandeau, E., Grasset, J., Defo, M. A., Campbell, P. G. C., Couture, P., & Bernatchez, L. (2016). Transcriptional response of yellow perch to changes in ambient metal concentrations—a reciprocal field transplantation experiment. *Aquatic Toxicology*, **173**, 132–142.

Bourret, V., Couture, P., Campbell, P. G. C., & Bernatchez, L. (2008). Evolutionary ecotoxicology of wild yellow perch (*Perca flavescens*) populations chronically exposed to a polymetallic gradient. *Aquatic Toxicology*, **86**(1), 76–90.

Boutron, C. F., Candelone, J.-P., & Hong, S. (1995). Greenland snow and ice cores: unique archives of large-scale pollution of the troposphere of the Northern Hemisphere by lead and other heavy metals. *Science of the total environment*, **160–161**, 233–241.

Boyle, D., Brix, K. V, Amlund, H., Lundebye, A.-K., Hogstrand, C., & Bury, N. R. (2008). Natural arsenic contaminated diets perturb reproduction in fish. *Environmental Science & Technology*, **42**(14), 5354–60.

Bradley, R. W., & Morris, J. R. (1986). Heavy metals in fish from a series of metal-contaminated lakes near Sudbury, Ontario. *Water, Air, & Soil Pollution*, **27**(3–4), 341–354.

Bremner, I. (1987). Involvement of metallothionein in the hepatic metabolism of copper. *The Journal of Nutrition*, **117**(1), 19–29.

Bridges, C. C., & Zalups, R. K. (2005). Molecular and ionic mimicry and the transport of toxic metals. *Toxicology and Applied Pharmacology*, **204**(3), 274–308.

Brinkman, S. F., & Hansen, D. L. (2007). Toxicity of cadmium to early life stages of brown trout (*Salmo trutta*) at multiple water hardnesses. *Environmental Toxicology and Chemistry*, **26**(8), 1666–1671.

Brinkman, S. F., & Woodling, J. D. (2014). Acclimation and deacclimation of brown trout (*Salmo trutta*) to zinc and copper singly and in combination with

cadmium or copper. *Archives of Environmental Contamination and Toxicology*, **67**(2), 214–223.

Brown, B. (1977). Effects of mine drainage on the River Hayle, Cornwall a) factors affecting concentrations of copper, zinc and iron in water, sediments and dominant invertebrate fauna. *Hydrobiologia*, **52**, 2–3.

Brown, B. E. (1976). Observations on the tolerance of the isopod *Asellus meridianus* to copper and lead. *Water Research*, **10**(6), 555–559.

Bury, N. R., & Wood, C. M. (1999). Mechanism of branchial apical silver uptake by rainbow trout is via the proton-coupled Na^+ channel. *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology*, **277**(5), R1385-91.

Bury, N. R., Walker, P. A., & Glover, C. N. (2003). Nutritive metal uptake in teleost fish. *Journal of Experimental Biology*, **206**(1), 11-23.

Bury, N., Galvez, F., & Wood, C. (1999). Effects of chloride, calcium, and dissolved organic carbon on silver toxicity: Comparison between rainbow trout and fathead minnows. *Environmental Toxicology and Chemistry,* **18**(1), 56–62.

Butler, M., Haskew, A. E. J., & Young, M. M. (1980). Copper tolerance in the green alga, *Chlorella vulgaris*. *Plant, Cell and Environment*, **3**(2), 119–126.

Calabrese, E. J., & Baldwin, L. A. (1997). The dose determines the stimulation (and poison): development of a chemical hormesis database. *International Journal of Toxicology*, **16**(6), 545–559.

Campbell, P. G. C., Giguère, A., Bonneris, E., & Hare, L. (2005). Cadmium-handling strategies in two chronically exposed indigenous freshwater organisms—the yellow perch (*Perca flavescens*) and the floater mollusc (*Pyganodon grandis*). *Aquatic Toxicology*, **72**(1–2), 83–97.

Campbell, P. G. C., Kraemer, L. D., Giguère, A., Hare, L., & Hontela, A. (2008). Subcellular distribution of cadmium and nickel in chronically exposed wild fish: inferences regarding metal detoxification strategies and implications for setting water quality guidelines for dissolved metals. *Human and Ecological Risk Assessment: An International Journal*, **14**(2), 290–316.

Candelone, J.-P., Hong, S., Pellone, C., & Boutron, C. F. (1995). Post-Industrial Revolution changes in large-scale atmospheric pollution of the northern hemisphere by heavy metals as documented in central Greenland snow and ice. *Journal of Geophysical Research*, **100**(D8), 16605-16616.

Canli, M., & Atli, G. (2003). The relationships between heavy metal (Cd, Cr, Cu, Fe, Pb, Zn) levels and the size of six Mediterranean fish species. *Environmental Pollution*, **121**(1), 129–136.

Ceyhun, S. B., Şentürk, M., Yerlikaya, E., Erdoğan, O., Küfrevioğlu, Ö. İ., & Ekinci, D. (2011). Purification and characterization of carbonic anhydrase

from the teleost fish *Dicentrarchus labrax* (European seabass) liver and toxicological effects of metals on enzyme activity. *Environmental Toxicology and Pharmacology*, **32**(1), 69–74.

Chakoumakos, C., Russo, R. C., & Thurston, R. V. (1979). Toxicity of copper to cutthroat trout (*Salmo clarki*) under different conditions of alkalinity, pH, and hardness. *Environmental Science & Technology*, **13**(2), 213–219.

Chowdhury, M. J., Bucking, C., & Wood, C. M. (2008). Pre-exposure to waterborne nickel downregulates gastrointestinal nickel uptake in rainbow trout: indirect evidence for nickel essentiality. *Environmental Science & Technology*, **42**(4), 1359–64.

Chowdhury, M. J., Girgis, M., & Wood, C. M. (2016). Revisiting the mechanisms of copper toxicity to rainbow trout: Time course, influence of calcium, unidirectional Na^+ fluxes, and branchial Na^+, K^+ ATPase and V-type H^+ ATPase activities. *Aquatic Toxicology*, **177**, 51–62.

Chowdhury, M. J., Grosell, M., McDonald, D., & Wood, C. (2003). Plasma clearance of cadmium and zinc in non-acclimated and metal-acclimated trout. *Aquatic Toxicology*, **64**(3), 259–275.

Clemow, Y.H., & Wilkie, M. P. (2015). Effects of Pb plus Cd mixtures on toxicity, and internal electrolyte and osmotic balance in the rainbow trout (*Oncorhynchus mykiss*). *Aquatic Toxicology*, **161**, 176–188.

Crump, K. L., & Trudeau, V. L. (2009). Mercury-induced reproductive impairment in fish. *Environmental Toxicology and Chemistry*, **28**(5), 895-907.

da Silva, E. S., Abril, S. I. M., Zanette, J., & Bianchini, A. (2014). Salinity-dependent copper accumulation in the guppy *Poecilia vivipara* is associated with CTR1 and ATP7B transcriptional regulation. *Aquatic Toxicology*, **152**, 300–307.

Dalzell, D. J. B., & Macfarlane, N. A. A. (1999). The toxicity of iron to brown trout and effects on the gills: a comparison of two grades of iron sulphate. *Journal of Fish Biology*, **55**(2), 301–315.

Dang, F., Zhong, H., & Wang, W.-X. (2009). Copper uptake kinetics and regulation in a marine fish after waterborne copper acclimation. *Aquatic Toxicology*, **94**(3), 238–244.

Dang, Z., Lock, R. A., Flik, G., & Wendlelaar Bonga, S. E. (1999). Metallothionein response in gills of *Oreochromis mossambicus* exposed to copper in fresh water. *The American Journal of Physiology*, **277**(1 Pt 2), R320-31.

De Boeck, G., Ngo, T. T. H., Van Campenhout, K., & Blust, R. (2003). Differential metallothionein induction patterns in three freshwater fish during sublethal copper exposure. *Aquatic Toxicology*, **65**(4), 413–424.

De Smet, H., & Blust, R. (2001). Stress responses and changes in protein metabolism in Carp *Cyprinus carpio* during cadmium exposure. *Ecotoxicology and Environmental Safety*, **48**(3), 255–262.

De Smet, H., De Wachter, B., Lobinski, R., & Blust, R. (2001). Dynamics of (Cd,Zn)-metallothioneins in gills, liver and kidney of common carp *Cyprinus carpio* during cadmium exposure. *Aquatic Toxicology*, **52**(3–4), 269–281.

Dragun, Z., Filipović Marijić, V., Kapetanović, D., Valić, D., Vardić Smrzlić, I., Krasnići, N., *et al.* (2013). Assessment of general condition of fish inhabiting a moderately contaminated aquatic environment. *Environmental Science and Pollution Research*, **20**(7), 4954–4968.

Dudley, R. E., Gammal, L. M., & Klaassen, C. D. (1985). Cadmium-induced hepatic and renal injury in chronically exposed rats: Likely role of hepatic cadmium-metallothionein in nephrotoxicity. *Toxicology and Applied Pharmacology*, **77**(3), 414–426.

Dural, M., Göksu, M. Z. L., & Özak, A. A. (2007). Investigation of heavy metal levels in economically important fish species captured from the Tuzla lagoon. *Food Chemistry*, **102**(1), 415–421.

Durrant, C. J., Stevens, J. R., Hogstrand, C., & Bury, N. R. (2011). The effect of metal pollution on the population genetic structure of brown trout (*Salmo trutta* L.) residing in the River Hayle, Cornwall, UK. *Environmental*, **159**(12), 3595–603.

Eroglu, A., Dogan, Z., Kanak, E. G., Atli, G., & Canli, M. (2015). Effects of heavy metals (Cd, Cu, Cr, Pb, Zn) on fish glutathione metabolism. *Environmental Science and Pollution Research*, **22**(5), 3229–3237.

Evans, D. H. (1987). The fish gill: site of action and model for toxic effects of environmental pollutants. *Environmental Health Perspectives*, **71**, 47–58.

Everall, N. C., Macfarlane, N. A. A., & Sedgwick, R. W. (1989). The effects of water hardness upon the uptake, accumulation and excretion of zinc in the brown trout, *Salmo trutta* L. *Journal of Fish Biology*, **35**(6), 881–892.

Facey, D. E., Blazer, V. S., Gasper, M. M., & Turcotte, C. L. (2005). Using fish biomarkers to monitor improvements in environmental quality. *Journal of Aquatic Animal Health*, **17**(3), 263–266.

Fernandes, M. N., Paulino, M. G., Sakuragui, M. M., Ramos, C. A., Pereira, C. D. S., & Sadauskas-Henrique, H. (2013). Organochlorines and metals induce changes in the mitochondria-rich cells of fish gills: An integrative field study involving chemical, biochemical and morphological analyses. *Aquatic Toxicology*, **126**, 180–190.

Förstner, U., & Wittmann, G. T. W. (1981). *Metal Pollution in the Aquatic Environment*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Foster, P. L. (1982). Species associations and metal contents of algae from rivers polluted by heavy metals. *Freshwater Biology*, **12**(1), 17–39.

Garcia-Santos, S., Fontaínhas-Fernandes, A., & Wilson, J. M. (2006). Cadmium tolerance in the Nile tilapia (*Oreochromis niloticus*) following acute exposure: Assessment of some ionoregulatory parameters. *Environmental Toxicology*, **21**(1), 33–46.

Giguere, A., Campbell, P., Hare, L., & Couture, P. (2006). Sub-cellular partitioning of cadmium, copper, nickel and zinc in indigenous yellow perch (*Perca flavescens*) sampled along a polymetallic gradient. *Aquatic Toxicology*, **77**(2), 178–189.

Gill, T. S., & Epple, A. (1993). Stress-related changes in the hematological profile of the American eel (*Anguilla rostrata*). *Ecotoxicology and Environmental Safety*, **25**(2), 227–235.

Gill, T. S., Bianchi, C. P., & Epple, A. (1992). Trace metal (Cu and Zn) adaptation of organ systems of the American eel, *Anguilla rostrata*, to external concentrations of cadmium. *Comparative Biochemistry and Physiology. C, Comparative Pharmacology and Toxicology*, **102**(3), 361–71.

Gimbert, F., Geffard, A., Guédron, S., Dominik, J., & Ferrari, B. J. D. (2016). Mercury tissue residue approach in *Chironomus riparius*: Involvement of toxicokinetics and comparison of subcellular fractionation methods. *Aquatic Toxicology*, **171**, 1–8.

Goss, G., Gilmour, K., Hawkings, G., Brumbach, J. H., Huynh, M., & Galvez, F. (2011). Mechanism of sodium uptake in PNA negative MR cells from rainbow trout, *Oncorhynchus mykiss* as revealed by silver and copper inhibition. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, **159**(3), 234–241.

Gower, A. M., Myers, G., Kent, M., & Foulkes, M. E. (1994). Relationships between macroinvertebrate communities and environmental variables in metal-contaminated streams in south-west England. *Freshwater Biology*, **32**(1), 199–221.

Grosell, M., Hogstrand, C., & Wood, C. M. (1997). Cu uptake and turnover in both Cu-acclimated and non-acclimated rainbow trout (*Oncorhynchus mykiss*). *Aquatic Toxicology*, **38**(4), 257–276.

Grosell, M. (2012). Copper. In: C.M. Wood, A.P. Farrell & C.J. Brauner (eds.), *Homeostasis and toxicology of essential metals*, Volume 31A. Cambridge, MA: Elsevier Academic Press, pp. 54-110.

Grosell, M. H., Hogstrand, C., & Wood, C. M. (1998). Renal Cu and Na excretion and hepatic Cu metabolism in both Cu acclimated and non-acclimated rainbow trout (*Oncorhynchus mykiss*). *Aquatic Toxicology*, **40**(2-3), 275–291.

Grosell, M., & Wood, C. M. (2002). Copper uptake across rainbow trout gills: mechanisms of apical entry. *The Journal of Experimental Biology*, **205**(8), 1179–88.

Grosell, M., McDonald, M. ., Walsh, P., & Wood, C. (2004). Effects of prolonged copper exposure in the marine gulf toadfish (Opsanus beta) II: copper accumulation, drinking rate and Na^+/K^+-ATPase activity in osmoregulatory tissues. *Aquatic Toxicology*, **68**(3), 263–275.

Grosell, M., McGeer, J. C., & Wood, C. M. (2001). Plasma copper clearance and biliary copper excretion are stimulated in copper-acclimated trout. *American Journal of Physiology. Regulatory, Integrative and Comparative Physiology*, **280**(3), R796-806.

Grosell, M., Nielsen, C., & Bianchini, A. (2002). Sodium turnover rate determines sensitivity to acute copper and silver exposure in freshwater animals. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology* **133**(1-2), 287-303.

Hansen, B. H., Garmo, Ø. A., Olsvik, P. A., & Andersen, R. A. (2007). Gill metal binding and stress gene transcription in brown trout (*Salmo trutta*) exposed to metal environments: the effect of pre-exposure in natural populations. *Environmental Toxicology and Chemistry*, **26**(5), 944.

Hansen, B. H., Rømma, S., Garmo, Ø. A., Olsvik, P. A., & Andersen, R. A. (2006a). Antioxidative stress proteins and their gene expression in brown trout (*Salmo trutta*) from three rivers with different heavy metal levels. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, **143**(3), 263–274.

Hansen, J., Lipton, J., Welsh, P., Morris, J., Cacela, D., & Suedkamp, M. (2002). Relationship between exposure duration, tissue residues, growth, and mortality in rainbow trout (*Oncorhynchus mykiss*) juveniles sub-chronically exposed to copper. *Aquatic Toxicology*, **58**(3–4), 175–188.

Heerden, D. van, Vosloo, A., & Nikinmaa, M. (2004). Effects of short-term copper exposure on gill structure, metallothionein and hypoxia-inducible factor-1α (HIF-1α) levels in rainbow trout (*Oncorhynchus mykiss*). *Aquatic Toxicology*, **69**(3), 271–280.

Heerden, D. van, Vosloo, A., & Nikinmaa, M. (2004). Effects of short-term copper exposure on gill structure, metallothionein and hypoxia-inducible factor-1α (HIF-1α) levels in rainbow trout (*Oncorhynchus mykiss*). *Aquatic Toxicology*, **69**(3), 271–280.

Hogstrand, C. (2012). Zinc. In: C.M. Wood, A.P. Farrell & C.J. Brauner (eds.), *Homeostasis and toxicology of essential metals*, Volume 31A. Cambridge, MA: Elsevier Academic Press, pp. 136-184.

Hogstrand, C. & Wood, CM. (1996). The physiology and toxicology of zinc in fish. In Taylor, E.M. (ed.) Toxicology of aquatic pollution: physiological, molecular and cellular approaches. Cambridge: Cambridge University Press, pp. 61-85

Hogstrand, C., & Haux, C. (1991). Binding and detoxification of heavy metals in lower vertebrates with reference to metallothionein. *Comparative*

Biochemistry and Physiology Part C: Comparative Pharmacology, **100**(1–2), 137–141.

Hogstrand, C., Lithner, G., & Haux, C. (1991). The importance of metallothionein for the accumulation of copper, zinc and cadmium in environmentally exposed perch, *Perca fluviatilis*. *Pharmacology & Toxicology*, **68**(6), 492–501.

Holcombe, G. W., Benoit, D. A., & Leonard, E. N. (1979). Long-term effects of zinc exposures on brook trout (*Salvelinus fontinalis*). *Transactions of the American Fisheries Society*, **108**(1), 76–87.

Holeton, G. F., & Randall, D. J. (1967). The effect of hypoxia upon the partial pressure of gases in the blood and water afferent and efferent to the gills of rainbow trout. *The Journal of Experimental Biology*, **46**(2), 317–27.

Hollis, L., Hogstrand, C., & Wood, C. M. (2001). Tissue-specific cadmium accumulation, metallothionein induction, and tissue zinc and copper levels during chronic sublethal cadmium exposure in juvenile rainbow trout. *Archives of Environmental Contamination and Toxicology*, **41**(4), 468–474.

Hollis, L., McGeer, J. C., McDonald, D. G., & Wood, C. M. (1999). Cadmium accumulation, gill Cd binding, acclimation, and physiological effects during long term sublethal Cd exposure in rainbow trout. *Aquatic Toxicology*, **46**(2), 101–119.

Hoseini, S. M., Rajabiesterabadi, H., & Kordrostami, S. (2016). Chronic exposure of *Rutilus rutilus caspicus* fingerlings to ambient copper. *Toxicology and Industrial Health*, **32**(2), 375–383.

Hughes, C.M. (1984). General anatomy of the gills. In: W.S. Hoar, D.J. Randall (eds.) *Fish Physiology*. New York: Academic Press. pp. 1–72.

Hung, C. C., Nawata, C. M., Wood, C. M., & Wright, P. A. (2008). Rhesus glycoprotein and urea transporter genes are expressed in early stages of development of rainbow trout (*Oncorhynchus mykiss*). *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, **309A**(5), 262–268.

Ito, Y., Kobayashi, S., Nakamura, N., Miyagi, H., Esaki, M., Hoshijima, K., & Hirose, S. (2013). Close association of carbonic anhydrase (CA2a and CA15a), Na^+/H^+ exchanger (Nhe3b), and ammonia transporter Rhcg1 in zebrafish ionocytes responsible for Na^+ uptake. *Frontiers in Physiology*, **4**, 59.

Ivancic, I., & Degobbis, D. (1984). An optimal manual procedure for ammonia analysis in natural waters by the indophenol blue method. *Water Research*, **18**(9), 1143–1147.

Kamunde, C. (2009). Early subcellular partitioning of cadmium in gill and liver of rainbow trout (*Oncorhynchus mykiss*) following low-to-near-lethal waterborne cadmium exposure. *Aquatic Toxicology*, **91**(4), 291–301.

Kamunde, C., Grosell, M., Higgs, D., & Wood, C. M. (2002a). Copper metabolism in actively growing rainbow trout (*Oncorhynchus mykiss*): interactions between dietary and waterborne copper uptake. *The Journal of Experimental Biology*, **205**(2), 279–90.

Kamunde, C., Clayton, C., & Wood, C. M. (2002b). Waterborne vs. dietary copper uptake in rainbow trout and the effects of previous waterborne copper exposure. *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology*, **283**(1), R69–R78.

Kawatsu, H. (1972). Studies on the anaemia of fish: V. Dietary iron deficient anaemia in brook trout, *Salvelinus fontinalis*. *Bull Freshwater Fish Res Lab.* **22**(1), 59-67.

Khan, F. R., Irving, J. R., Bury, N. R., & Hogstrand, C. (2011). Differential tolerance of two *Gammarus pulex* populations transplanted from different metallogenic regions to a polymetal gradient. *Aquatic Toxicology*, **102**(1), 95–103.

Khan, R. A. (2003). Health of flatfish from localities in Placentia Bay, Newfoundland, contaminated with petroleum and PCBs. *Archives of Environmental Contamination and Toxicology*, **44**(4), 485–492.

Klaassen, C. D., Liu, J., & Choudhuri, S. (1999). Metallothionein: an intracellular protein to protect against cadmium toxicity. *Annual Review of Pharmacology and Toxicology*, **39**(1), 267–294.

Klerks, P. L., & Bartholomew, P. R. (1991). Cadmium accumulation and detoxification in a Cd-resistant population of the oligochaete *Limnodrilus hoffmeisteri*. *Aquatic Toxicology*, **19**(2), 97–112.

Köck, G., & Bucher, F. (1997). Accumulation of zinc in rainbow trout (*Oncorhynchus mykiss*) after waterborne and dietary exposure. *Bulletin of Environmental Contamination and Toxicology*, **58**(2), 305–10.

Kraemer, L. D., Campbell, P. G. C., & Hare, L. (2005). Dynamics of Cd, Cu and Zn accumulation in organs and sub-cellular fractions in field transplanted juvenile yellow perch (*Perca flavescens*). *Environmental Pollution*, **138**(2), 324–337.

Lange, A., Ausseil, O., & Segner, H. (2002). Alterations of tissue glutathione levels and metallothionein mRNA in rainbow trout during single and combined exposure to cadmium and zinc. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, **131**(3), 231–243.

Lanno, R. P., Hicks, B., & Hilton, J. W. (1987). Histological observations on intrahepatocytic copper-containing granules in rainbow trout reared on diets containing elevated levels of copper. *Aquatic Toxicology*, **10**(5–6), 251–263.

Lapointe, D., & Couture, P. (2009). Influence of the route of exposure on the accumulation and subcellular distribution of nickel and thallium in juvenile

fathead minnows (*Pimephales promelas*). *Archives of Environmental Contamination and Toxicology*, **57**(3), 571–580.

Lappivaara, J., Nikinmaa, M., & Tuurala, H. (1995). Arterial oxygen tension and the structure of the secondary lamellae of the gills in rainbow trout (*Oncorhynchus mykiss*) after acute exposure to zinc and during recovery. *Aquatic Toxicology*, **32**(4), 321–331.

Larsen, B. K., Pörtner, H.-O., & Jensen, F. B. (1997). Extra- and intracellular acid-base balance and ionic regulation in cod (*Gadus morhua*) during combined and isolated exposures to hypercapnia and copper. *Marine Biology*, **128**(2), 337–346.

Larsson, Å., Haux, C., & Sjöbeck, M.-L. (1985). Fish physiology and metal pollution: results and experiences from laboratory and field studies. *Ecotoxicology and Environmental Safety*, **9**(3), 250–281.

Lauff, R. F., & Wood, C. M. (1996). Respiratory gas exchange, nitrogenous waste excretion, and fuel usage during starvation in juvenile rainbow trout, *Oncorhynchus mykiss*. *Journal of Comparative Physiology. B, Biochemical, Systemic, and Environmental Physiology*, **165**(7), 542–51.

Laughlin, R. B., Ng, J., & Guard, H. E. (1981). Hormesis: A response to low environmental concentrations of petroleum hydrocarbons. *Science*, **211**(4483), 705–707.

Laurén, D. J., & McDonald, D. G. (1985). Effects of copper on branchial ionoregulation in the rainbow trout, *Salmo gairdneri*. *Journal of Comparative Physiology B*, **155**(5), 635–644.

Laurén, D. J., & McDonald, D. G. (1987). Acclimation to copper by rainbow trout, *Salmo gairdneri* : Biochemistry. *Canadian Journal of Fisheries and Aquatic Sciences*, **44**(1), 105–111.

Lefevre, S., Jensen, F. B., Huong, D. T. T., Wang, T., Phuong, N. T., & Bayley, M. (2012). Haematological and ion regulatory effects of nitrite in the air-breathing snakehead fish *Channa striata*. *Aquatic Toxicology*, **118–119**, 48–53.

Lim, M. Y.-T., Zimmer, A. M., & Wood, C. M. (2015). Acute exposure to waterborne copper inhibits both the excretion and uptake of ammonia in freshwater rainbow trout (*Oncorhynchus mykiss*). *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, **168**, 48–54.

Maes, G. E., Raeymaekers, J. A. M., Pampoulie, C., Seynaeve, a, Goemans, G., Belpaire, C., & Volckaert, F. A. M. (2005). The catadromous European eel *Anguilla anguilla* (L.) as a model for freshwater evolutionary ecotoxicology: relationship between heavy metal bioaccumulation, condition and genetic variability. *Aquatic Toxicology*, **73**(1), 99–114.

Majewski, H. (1981). Cardiovascular-respiratory responses of rainbow trout (*Salmo gairdneri*) during chronic exposure to sublethal concentrations of cadmium. *Water Research*, **15**(10), 1211–1217.

Mason, A.Z., & Jenkins, K.D. (1995). Metal detoxification in aquatic organisms. In: A. Tessier, D.R. Turner (eds.), *Metal speciation and bioavailability in aquatic systems*. Chichester, UK: John Wiley & Sons, pp. 479-608.

McDonald *et al.* 1989. The combined effects of pH and trace metals on fish ionoregulation. In Morris, R., Taylor, E.W., Brown, D.J.A. and Brown, J.A. (eds). *Acid Toxicity and Aquatic Animals.* Cambridge: Cambridge University Press, pp. 221-242.

McDonald, D. G., & Wood, C. M. (1993). Branchial mechanisms of acclimation to metals in freshwater fish. In: J.C. Rankin & F.B. Jensen (eds.) *Fish Ecophysiology*. Dordrecht: Springer Netherlands, pp. 297–321.

McGeer, J. C., Nadella, S., Alsop, D. H., Hollis, L., Taylor, L. N., McDonald, D. G., & Wood, C. M. (2007). Influence of acclimation and cross-acclimation of metals on acute Cd toxicity and Cd uptake and distribution in rainbow trout (*Oncorhynchus mykiss*). *Aquatic Toxicology,* **84**(2), 190–7.

McGeer, J. C., Szebedinszky, C., Gordon McDonald, D., & Wood, C. M. (2000). Effects of chronic sublethal exposure to waterborne Cu, Cd or Zn in rainbow trout II: tissue specific metal accumulation. *Aquatic Toxicology*, **50**(3), 245–256.

McGeer, J.C., Niyogi, S., & Smith, D.S. (2012). Cadmium. In: C.M. Wood, A.P. Farrell & C.J. Brauner (eds.), *Homeostasis and toxicology of non-essential metals*, Volume 31B. Cambridge, MA: Elsevier Academic Press, pp. 126-168.

Meyer, J. N., & Di Giulio, R. T. (2003). Heritable adaptation and fitness costs in killifish (*Fundulus heteroclitus*) inhabiting a polluted estuary. *Ecological Applications*, **13***(*2), 490–503.

Minghetti, M., Leaver, M. J., & George, S. G. (2010). Multiple Cu-ATPase genes are differentially expressed and transcriptionally regulated by Cu exposure in sea bream, *Sparus aurata*. *Aquatic Toxicology*, **97**(1), 23–33.

Minghetti, M., Leaver, M. J., Carpenè, E., & George, S. G. (2008). Copper transporter 1, metallothionein and glutathione reductase genes are differentially expressed in tissues of sea bream (*Sparus aurata*) after exposure to dietary or waterborne copper. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, **147**(4), 450–459.

Minghetti, M., Schnell, S., Chadwick, M. A., Hogstrand, C., & Bury, N. R. (2014). A primary FIsh Gill Cell System (FIGCS) for environmental monitoring of river waters. *Aquatic Toxicology*, **154**, 184–192.

Monna, F., Camizuli, E., Revelli, P., Biville, C., Thomas, C., Losno, R., *et al.* (2011). Wild brown trout affected by historical mining in the Cévennes

National Park, France. *Environmental Science & Technology*, **45**(16), 6823–6830.

Monteiro, S. M., Mancera, J. M., Fontaínhas-Fernandes, A., & Sousa, M. (2005). Copper induced alterations of biochemical parameters in the gill and plasma of *Oreochromis niloticus*. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, **141**(4), 375–383.

Nakada, T., Westhoff, C. M., Kato, A., & Hirose, S. (2007). Ammonia secretion from fish gill depends on a set of Rh glycoproteins. *FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology*, **21**(4), 1067–74.

Nawata, C. M., Hung, C. C. Y., Tsui, T. K. N., Wilson, J. M., Wright, P. A., & Wood, C. M. (2007). Ammonia excretion in rainbow trout (*Oncorhynchus mykiss*): evidence for Rh glycoprotein and H^+-ATPase involvement. *Physiological Genomics*, **31**(3), 463–474.

Nonnotte, G., & Truchot, J.-P. (1990). Time course of extracellular acid-base adjustments under hypo- or hyperosmotic conditions in the euryhaline fish *Platichthys flesus*. *Journal of Fish Biology*, **36**(2), 181–190.

Nriagu, J. (1990). Human influence on the global cycling of trace metals. *Global and Planetary Change*, **2**(1-2), 113–120.

Nriagu, J. O. (1996). A history of global metal pollution. *Science*, **272**(5259), 223.

Ownby, D. R., Newman, M. C., Mulvey, M., Vogelbein, W. K., Unger, M. A., & Arzayus, L. F. (2002). Fish (*Fundulus heteroclitus*) populations with different exposure histories differ in tolerance of creosote-contaminated sediments. *Environmental Toxicology and Chemistry*, **21**(9), 1897–1902.

Pain, D. ., Sánchez, A., & Meharg, A. (1998). The Doñana ecological disaster: contamination of a world heritage estuarine marsh ecosystem with acidified pyrite mine waste. *Science of the Total Environment*, **222**(1), 45–54.

Palermo, F. F., Risso, W. E., Simonato, J. D., & Martinez, C. B. R. (2015). Bioaccumulation of nickel and its biochemical and genotoxic effects on juveniles of the neotropical fish *Prochilodus lineatus*. *Ecotoxicology and Environmental Safety*, **116**, 19–28.

Pane, E. F., Bucking, C., Patel, M., & Wood, C. M. (2005). Renal function in the freshwater rainbow trout (*Oncorhynchus mykiss*) following acute and prolonged exposure to waterborne nickel. *Aquatic Toxicology*, **72**(1–2), 119–133.

Pane, E. F., Glover, C. N., Patel, M., & Wood, C. M. (2006). Characterization of Ni transport into brush border membrane vesicles (BBMVs) isolated from the kidney of the freshwater rainbow trout (*Oncorhynchus mykiss*). *Biochimica et Biophysica Acta (BBA) - Biomembranes*, **1758**(1), 74–84.

Pane, E. F., Haque, A., Goss, G. G., & Wood, C. M. (2004). The physiological consequences of exposure to chronic, sublethal waterborne nickel in rainbow trout (*Oncorhynchus mykiss*): exercise vs resting physiology. *The Journal of Experimental Biology*, **207**(Pt 7), 1249–61.

Pane, E. F., Patel, M., & Wood, C. M. (2006). Chronic, sublethal nickel acclimation alters the diffusive properties of renal brush border membrane vesicles (BBMVs) prepared from the freshwater rainbow trout. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, **143**(1), 78–85.

Pane, E. F., Richards, J. G., & Wood, C. M. (2003). Acute waterborne nickel toxicity in the rainbow trout (*Oncorhynchus mykiss*) occurs by a respiratory rather than ionoregulatory mechanism. *Aquatic Toxicology*, **63**(1), 65–82.

Paris, J. R., King, R. A., & Stevens, J. R. (2015). Human mining activity across the ages determines the genetic structure of modern brown trout (*Salmo trutta* L.) populations. *Evolutionary Applications*, **8**(6), 573–585.

Peterson, M. S., & Gilmore, R. G. (1988). Hematocrit, osmolality, and ion concentration in fishes: consideration of circadian patterns in the experimental design. *Journal of Experimental Marine Biology and Ecology*, **121**(1), 73–78.

Pirrie, D., Beer, A.J., & Camm, G.S. (1999). Early diagnetic sulphide minerals in the Hayle Estuary, Cornwall, *Geoscience in South-West England*, **9**(4):325-332.

Pirrie, D., Power, M. R., Rollinson, G., Camm, G. S., Hughes, S. H., Butcher, A. R., & Hughes, P. (2003). The spatial distribution and source of arsenic, copper, tin and zinc within the surface sediments of the Fal Estuary, Cornwall, UK. *Sedimentology*, **50**(3), 579–595.

Pirrie, D., Power, M. R., Wheeler, P. D., Cundy, A., Bridges, C., & Davey, G. (2002). Geochemical signature of historical mining: Fowey Estuary, Cornwall, UK. *Journal of Geochemical Exploration*, **76**(1), 31–43.

Pirrie, D., Sear, L. G., Hughes, S. H., & Camm, G. S. (1997). Mineralogical and geochemical signature of mine waste contamination, Tresillian River, Fal Estuary, Cornwall, UK. *Environmental Geology*, **29**(1–2), 58–65.

Playle, R. C. (2004). Using multiple metal–gill binding models and the toxic unit concept to help reconcile multiple-metal toxicity results. *Aquatic Toxicology*, **67**(4), 359–370.

Playle, R. C., Dixon, D. G., & Burnison, K. (1993). Copper and cadmium binding to fish gills: Estimates of metal–gill stability constants and modelling of metal accumulation. *Canadian Journal of Fisheries and Aquatic Sciences*, **50**(12), 2678–2687.

Poleksic, V., Lenhardt, M., Jaric, I., Djordjevic, D., Gacic, Z., Cvijanovic, G., & Raskovic, B. (2010). Liver, gills, and skin histopathology and heavy metal

content of the Danube sterlet (*Acipenser ruthenus*). *Environmental Toxicology and Chemistry*, **29**(3), 515–521.

Ptashynski, M. D., Pedlar, R. M., Evans, R. E., Wautier, K. G., Baron, C. L., & Klaverkamp, J. F. (2001). Accumulation, distribution and toxicology of dietary nickel in lake whitefish (*Coregonus clupeaformis*) and lake trout (*Salvelinus namaycush*). *Comparative Biochemistry and Physiology. Toxicology & Pharmacology : CBP*, **130**(2), 145–62.

Pyle, G., & Couture, P. (2012). Nickel. In: C.M. Wood, A.P. Farrell & C.J. Brauner (eds.), *Homeostasis and toxicology of essential metals*, Volume 31A. Cambridge, MA: Elsevier Academic Press, pp. 254-281.

Qiu, A., Glover, C. N., & Hogstrand, C. (2007). Regulation of branchial zinc uptake by 1α,25-(OH)2D3 in rainbow trout and associated changes in expression of ZIP1 and ECaC. *Aquatic Toxicology*, **84**(2), 142–152.

Rana, S. V. S., & Singh, R. (1996). Species differences in glutathione-dependent enzymes in the liver and kidney of two fresh water fishes and their implications for cadmium toxicity. *Ichthyological Research*, **43**(3), 223–229.

Randall, D. ., & Tsui, T. K. (2002). Ammonia toxicity in fish. *Marine Pollution Bulletin*, **45**(1–12), 17–23.

Randi, A. S., Monserrat, J. M., Rodriguez, E. M., & Romano, L. A. (1996). Histopathological effects of cadmium on the gills of the freshwater fish, *Macropsobrycon uruguayanae* Eigenmann (Pisces, Atherinidae). *Journal of Fish Diseases*, **19**(4), 311–322.

Reynders, H., Van Campenhout, K., Bervoets, L., De Coen, W. M., & Blust, R. (2006). Dynamics of cadmium accumulation and effects in common carp (*Cyprinus carpio*) during simultaneous exposure to water and food (*Tubifex tubifex*). *Environmental Toxicology and Chemistry*, **25**(6), 1558-67.

Rios, F. S., Moraes, G., Oba, E. T., Fernandes, M. N., Donatti, L., Kalinin, A. L., & Rantin, F. T. (2006). Mobilization and recovery of energy stores in traíra, *Hoplias malabaricus* Bloch (Teleostei, Erythrinidae) during long-term starvation and after re-feeding. *Journal of Comparative Physiology B*, **176**(7), 721–728.

Roy, R., & Bhoite, S. (2016). Compensatory adjustment in chloride cells during salinity adaptation in mud crab, *Scylla serrata*. *Proceedings of the Zoological Society*, **69**(2), 242–248.

Ruane, N. M., Wendelaar Bonga, S. E., & Balm, P. H. M. (1999). Differences between rainbow trout and brown trout in the regulation of the pituitary–interrenal axis and physiological performance during confinement. *General and Comparative Endocrinology*, **115**(2), 210–219.

Safahieh, A., Hedayati, A., Savari, A., & Movahedinia, A. (2010). Experimental approaches of hematotoxic and immunotoxic effects of mercury chloride on

yellow fin sea bream (*Acanthopagrus latus*). *American-Eurasian Journal of Toxicological Sciences,* **2**(3), 169-176.

Santore, R. C., Mathew, R., Paquin, P. R., & DiToro, D. (2002). Application of the biotic ligand model to predicting zinc toxicity to rainbow trout, fathead minnow, and *Daphnia magna*. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, **133**(1–2), 271–285.

Scott, G. R., & Sloman, K. A. (2004). The effects of environmental pollutants on complex fish behaviour: integrating behavioural and physiological indicators of toxicity. *Aquatic Toxicology*, **68**(4), 369–392.

Shoemaker, C. A., Klesius, P. H., Lim, C., & Yildirim, M. (2003). Feed deprivation of channel catfish, *Ictalurus punctatus*, influences organosomatic indices, chemical composition and susceptibility to *Flavobacterium columnare*. *Journal of Fish Diseases*, **26**(9), 553–561.

Sinha, A. K., Kapotwe, M., Dabi, S. B., Montes, C. da S., Shrivastava, J., Blust, R., & Boeck, G. De. (2016). Differential modulation of ammonia excretion, Rhesus glycoproteins and ion-regulation in common carp (*Cyprinus carpio*) following individual and combined exposure to waterborne copper and ammonia. *Aquatic Toxicology*, **170**, 129–141.

Skidmore, J. F. (1964). Toxicity of zinc compounds to aquatic animals, with special reference to fish. *The Quarterly Review of Biology*, **39**(3), 227–248.

Skidmore, J. F. (1970). Respiration and osmoregulation in rainbow trout with gills damaged by zinc sulphate. *Journal of Experimental Biology*, **52**(2), 481-494.

Sloman, K. A., Gilmour, K. M., Taylor, A. C., & Metcalfe, N. B. (2000). Physiological effects of dominance hierarchies within groups of brown trout, *Salmo trutta*, held under simulated natural conditions. *Fish Physiology and Biochemistry*, **22**(1), 11–20.

Soivio, A., Westman, K., & Nyholm, K. (1974). Changes in haematocrit values in blood samples treated with and without oxygen: a comparative study with four salmonid species. *Journal of Fish Biology*, **6**(6), 763–769.

Spry, D. J., & Wood, C. M. (1985). Ion flux rates, acid–base status, and blood gases in rainbow trout, *Salmo gairdneri*, exposed to toxic zinc in natural soft water. *Canadian Journal of Fisheries and Aquatic Sciences*, **42**(8), 1332–1341.

Spry, D. J., Hodson, P. V., & Wood, C. M. (1988). Relative contributions of dietary and waterborne zinc in the rainbow trout, *Salmo gairdneri*. *Canadian Journal of Fisheries and Aquatic Sciences*, **45**(1), 32–41.

Staurnes, M., Sigholt, T., & Reite, O. B. (1984). Reduced carbonic anhydrase and Na^+−K^+-ATPase activity in gills of salmonids exposed to aluminium-containing acid water. *Experientia*, **40**(2), 226–227.

Suvetha, L., Ramesh, M., & Saravanan, M. (2010). Influence of cypermethrin toxicity on ionic regulation and gill Na^+/K^+-ATPase activity of a freshwater teleost fish *Cyprinus carpio*. *Environmental Toxicology and Pharmacology*, **29**(1), 44–49.

Tate-Boldt, E. K., & Kolok, A. S. (2008). Copper acclimation in juvenile fathead minnows: Is a cycle of branchial damage and repair necessary? *Aquatic Toxicology*, **87**(1), 13–18.

Tchounwou, P. B., Yedjou, C. G., Patlolla, A. K., & Sutton, D. J. (2012). Heavy metal toxicity and the environment. *EXS*, **101**, 133–64.

Thophon, S., Kruatrachue, M., Upatham, E. ., Pokethitiyook, P., Sahaphong, S., & Jaritkhuan, S. (2003). Histopathological alterations of white seabass, *Lates calcarifer*, in acute and subchronic cadmium exposure. *Environmental Pollution*, **121**(3), 307–320.

Todd, A. S., Brinkman, S., Wolf, R. E., Lamothe, P. J., Smith, K. S., & Ranville, J. F. (2009). An enriched stable-isotope approach to determine the gill–zinc binding properties of juvenile rainbow trout (*Oncorhynchus mykiss*) during acute zinc exposures in hard and soft waters. *Environmental Toxicology and Chemistry*, **28**(6), 1233-43.

Urbina, M. A., & Glover, C. N. (2015). Effect of salinity on osmoregulation, metabolism and nitrogen excretion in the amphidromous fish, inanga (*Galaxias maculatus*). *Journal of Experimental Marine Biology and Ecology*, **473**, 7–15.

Uren Webster, T. M., Bury, N., van Aerle, R., & Santos, E. M. (2013). Global transcriptome profiling reveals molecular mechanisms of metal tolerance in a chronically exposed wild population of brown trout. *Environmental Science & Technology*, **47**(15), 8869–77.

Van Campenhout, K., Bervoets, L., & Blust, R. (2003). Metallothionein concentrations in natural populations of gudgeon (*Gobio gobio*): Relationship with metal concentrations in tissues and environment. *Environmental Toxicology and Chemistry*, **22**(7), 1548–1555.

Van Campenhout, K., Infante, H. G., Adams, F., & Blust, R. (2004). Induction and binding of Cd, Cu, and Zn to metallothionein in carp (*Cyprinus carpio*) using HPLC-ICP-TOFMS. *Toxicological Sciences*, **80**(2), 276–287.

van Dyk, J. C., Cochrane, M. J., & Wagenaar, G. M. (2012). Liver histopathology of the sharptooth catfish *Clarias gariepinus* as a biomarker of aquatic pollution. *Chemosphere*, **87**(4), 301–311.

Van Hoof, F., & Van San, M. (1981). Analysis of copper, zinc, cadmium and chromium in fish tissues. A tool for detecting metal caused fish kills. *Chemosphere*, **10**(10), 1127–1135.

Vitale, A. M., Monserrat, J. M., Castilho, P., & Rodriguez, E. M. (1999). Inhibitory effects of cadmium on carbonic anhydrase activity and ionic

regulation of the estuarine crab *Chasmagnathus granulata* (Decapoda, Grapsidae). *Comparative Biochemistry and Physiology Part C: Pharmacology, Toxicology and Endocrinology*, **122**(1), 121–129.

Walsh, P.J. & Mommsen, T.P. 2001. Evolutionary considerations of nitrogen metabolism and excretion. In: Wright, P.A. & Anderson, P.M. (eds). *Nitrogen Excretion*. California, London: Academic Press, pp. 1-26.

Wang, T., Knudsen, P. K., Brauner, C. J., Busk, M., Vijayan, M. M., & Jensen, F. B. (1998). Copper exposure impairs intra- and extracellular acid-base regulation during hypercapnia in the fresh water rainbow trout (*Oncorhynchus mykiss*). *Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology*, **168**(8), 591–599.

Wang, W.-X., & Rainbow, P. S. (2013). Subcellular partitioning and the prediction of cadmium toxicity to aquatic organisms. *Environmental Chemistry*, **3**(6), 395-399.

Wang, W.-X., Yang, Y., Guo, X., He, M., Guo, F., & Ke, C. (2011). Copper and zinc contamination in oysters: Subcellular distribution and detoxification. *Environmental Toxicology and Chemistry*, **30**(8), 1767–1774.

Weber, C.I., 1993. Methods for measuring the acute toxicity of effluents to freshwater and marine organisms, 4th ed. *Environmental Monitoring and Support Laboratory*, U.S. Environmental Protection Agency, Cincinnati, OH. EPA/600/4-90/027F.

Wilson, R. W., & Taylor, E. W. (1993). The physiological responses of freshwater rainbow trout, *Oncorhynchus mykiss*, during acutely lethal copper exposure. *Journal of Comparative Physiology B*, **163**(1), 38–47.

Witeska, M. (2013). Erythrocytes in teleost fishes: a review. *Zoology and Ecology*, **23**(4), 275–281.

Witeska, M., Sarnowski, P., Ługowska, K., & Kowal, E. (2014). The effects of cadmium and copper on embryonic and larval development of ide *Leuciscus idus* L. *Fish Physiology and Biochemistry*, **40**(1), 151–163.

Wood, C. M. (1992). Flux measurements as indices of H^+ and metal effects on freshwater fish. *Aquatic Toxicology*, **22**(4), 239–263.

Wood, C. M., Farrell, A. P., & Brauner, C. J. (2012). *Homeostasis and toxicology of essential metals*. Academic Press.

Wood, C. M., Hogstrand, C., Galvez, F., & Munger, R. S. (1996). The physiology of waterborne silver toxicity in freshwater rainbow trout (*Oncorhynchus mykiss*) II. The effects of ionic Ag^+. *Aquatic Toxicology*, **35**(2), 93–109.

Wood, C. M., Nawata, C. M., Wilson, J. M., Laurent, P., Chevalier, C., Bergman, H. L. *et al.* (2013). Rh proteins and NH_4+-activated Na^+-ATPase in the

Magadi tilapia (*Alcolapia grahami*), a 100% ureotelic teleost fish. *Journal of Experimental Biology*, **216**(16), 2998–3007.

Wood, C. M., Playle, R. C., Simons, B. P., Goss, G. G., & McDonald, D. G. (1988). Blood gases, acid–base status, ions, and hematology in adult brook trout (*Salvelinus fontinalis*) under acid/aluminium exposure. *Canadian Journal of Fisheries and Aquatic Sciences*, **45**(9), 1575–1586.

Wright, P. A., & Wood, C. M. (2009). A new paradigm for ammonia excretion in aquatic animals: role of Rhesus (Rh) glycoproteins. *Journal of Experimental Biology*, **212**(15), 2303–2312.

Wright, P. A., Wood, C. M., & Wilson, J. M. (2014). Rh versus pH: the role of Rhesus glycoproteins in renal ammonia excretion during metabolic acidosis in a freshwater teleost fish. *Journal of Experimental Biology*, **217**(16), 2855–2865.

Yang, R., & Randall, D. J. (1996). Effects of Tetrachloroguaiacol (TeCG) on the osmoregulation of adult coho salmon (*Oncorhynchus kisutch*). *Bull. Environ. Contam. Toxicol*, **56**, 788–795.

Zheng, D., Feeney, G. P., Kille, P., & Hogstrand, C. (2008). Regulation of ZIP and ZnT zinc transporters in zebrafish gill: zinc repression of ZIP10 transcription by an intronic MRE cluster. *Physiological Genomics*, **34**(2), 205–214.

Zimmer, A. M., Barcarolli, I. F., Wood, C. M., & Bianchini, A. (2012). Waterborne copper exposure inhibits ammonia excretion and branchial carbonic anhydrase activity in euryhaline guppies acclimated to both fresh water and sea water. *Aquatic Toxicology*, **122**, 172–180.

Supporting Information

Brown trout display multi-faceted physiological mechanisms to cope with chronic metal toxicity in a metal-impacted river

This supporting information contains:

Table S1. Water chemistry features of synthetic freshwater during the experiment for the Fowey and Hayle populations.

Table S2. Metal concentration values (µg/L) of cadmium (Cd), copper (Cu), nickel (Ni) and zinc (Zn), measured in the synthetic freshwater for the exposures of each population: Fowey and Hayle.

Table S3. Sex ratios of the Fowey and Hayle populations across the sampling time points of the experiment.

Table S4. Measurements of the net fluxes in Na^+ and Cl^- of the Fowey and Hayle fish.

Table S5. Results for J_{net} flux of ammonia (NH_3) for the Fowey and Hayle populations across the 96-hour exposure.

Supporting Information Table 1. *Water chemistry features of synthetic freshwater during the experiment for the Fowey and Hayle populations*.

	Fowey			*Hayle*		
Time point	**pH**	**Temp. (°C)**	**Conductivity (µS/cm)**	**pH**	**Temp. (°C)**	**Conductivity (µS/cm)**
Depuration Day 1	7.31	15.5	270	7.29	15.5	270
Depuration Day 2	7.33	15.5	270	7.3	15.5	270
Depuration Day 3	7.31	15.6	270	7.32	15.3	270
Depuration Day 4	7.37	15.6	267	7.32	15.6	267
Depuration Day 5	7.26	15.5	265	7.3	15.5	269
Depuration Day 6	7.25	15.4	270	7.25	15.6	270
Depuration Day 7	7.28	15.5	270	7.26	15.5	268
Average	*7.3 ± 0.02*	*15.5 ± 0.03*	*269 ± 0.77*	*7.2 ± 0.01*	*15.5 ± 0.04*	*269 ± 0.46*
Exposure Day 1a	7.28	15.5	268	7.31	15.5	270
Exposure Day 1b	7.3	15.5	270	7.3	15.5	270
Exposure Day 2	7.3	15.4	270	7.31	15.6	270
Exposure Day 3	7.31	15.6	266	7.28	15.6	267
Exposure Day 4	7.3	15.5	268	7.32	15.5	265
Average	*7.3* ± 0.00	*15.5* ± 0.03	*268* ± 0.75	*7.3* ± 0.01	*15.5* ± 0.02	*268* ± 1.03

Daily measurements of pH, water temperature (°C) and conductivity (µS/cm) were assessed in order to verify the consistency of the synthetic water chemistry during depuration and during the exposure. Exposure 1a and 1b represent 1 hour prior to exposure, and 1 hour post exposure.

Supporting Information Table S2. *Metal concentration values (µg/L) of cadmium (Cd), copper (Cu), nickel (Ni) and zinc (Zn), measured in the synthetic freshwater for the exposures of each population: Fowey and Hayle.*

	Fowey				*Hayle*			
Time point	**Cd**	**Cu**	**Ni**	**Zn**	**Cd**	**Cu**	**Ni**	**Zn**
(-24 hours)	0.01	0.34	0.35	0.58	0.03	0.82	0.97	3.11
1 hour control	0.03	0.68	0.65	2.23	0.02	0.39	0.48	0.94
1 hour exposed	0.13	7.16	11.81	125.20	0.05	6.66	12.13	128.54
24 hours control	0.01	0.92	1.42	3.42	0.06	1.04	0.55	3.04
24 hours exposed	0.11	8.52	11.35	177.55	0.10	6.31	14.64	116.92
96 hours control	0.01	0.64	0.58	1.86	0.03	1.05	4.92	4.34
96 hours exposed	0.08	7.79	11.29	118.24	0.11	6.51	12.65	116.21
Average exposed	*0.106 ± 0.01*	*7.82 ± 0.39*	*11.48 ± 0.16*	*140.33 ± 18.72*	*0.09 ± 0.02*	*6.49 ± 0.10*	*13.14 ± 0.76*	*120.56 ± 4.00*
Average control	*0.02 ± 0.01*	*0.64 ± 0.12*	*0.75 ± 0.23*	*2.02 ± 0.59*	*0.04 ± 0.01*	*0.83 ± 0.15*	*1.73 ± 1.07*	*2.86 ± 0.70*
p-value	0.0016	0.0001	0.0001	0.0001	0.0587	0.0001	0.0001	0.0001

Statistically significant differences were observed between the metal concentration values of control and exposed tanks ($p<0.05$).

Supporting Information Table S3. *Sex ratios of the Fowey and Hayle populations across the sampling time points of the experiment.*

Time point	*Fowey Ratio M:F*	*Hayle Ratio M:F*
In-river	1.50	2.67
Depuration	1.17	1.40
24 hours control	1.17	0.83
24 hours exposed	0.75	0.50
96 hours control	1.00	1.33
96 hours exposed	0.78	1.33
TOTAL	1:1	1.18

Supporting Information Table S4. *Measurements of the net fluxes in Na^+ and Cl^- of the Fowey and Hayle fish.*

Na^+	*Time point*	*Fowey*	*Hayle*	*Significance*
	-24 hours	-1009 ± 1113	-293 ± 377	0.89 (p=0.376)
	0 hours	660 ± 475	54 ± 158	0.75 (p=0.454)
	1 hour	2540 ± 764	58 ± 161	3.08 (p=0.002)**
	8 hours	-258 ± 917	94 ± 441	0.45 (p=0.655)
	12 hours	-1130 ± 652	92 ± 650	1.56 (p=0.122)
	24 hours	344 ± 307	231 ± 299	0.14 (p=0.886)
	48 hours	521 ± 329	-56 ± 779	0.69 (p=0.489)
	72 hours	503 ± 219	11 ± 176	0.63 (p=0.532)
	96 hours	-244 ± 389	1708 ± 784	2.26 (p=0.025) *
Cl^-	*Time point*	*Fowey*	*Hayle*	*Significance*
	-24 hours	535 ± 331	-395 ± 477	1.33 (p=1.87)
	0 hours	649 ± 441	-478 ± 251	1.66 (p=0.089)
	1 hour	1467 ± 960	-1178 ± 605	3.7 (p<0.001)***
	8 hours	714 ± 400	963 ± 584	0.72 (p=0.471)
	12 hours	166 ± 616	-910 ± 548	1.5 (p=0.135)
	24 hours	328 ± 187	-12 ± 224	0.48 (p=0.636)
	48 hours	27 ± 288	405 ± 587	0.63 (p=0.529)
	72 hours	174 ± 186	-78 ± 137	0.35 (p=0.725)
	96 hours	-49 ± 169	1693 ± 885	2.85 (p=0.005)**

Measurements of the net Na^+ or Cl^- rate measured as µmol/kg/h using the equation of Wood (1992). Significance of a two-way ANOVA is shown, accepting significance at $p<0.05$, where $p \leq 0.05$ *; $p \leq 0.01$ ** and $p \leq 0.001$ ***

Supporting Information Table S5. *Results for J_{net} flux of ammonia (NH_3) for the Fowey and Hayle populations across the 96-hour exposure*.

Time point	*Fowey*	*Hayle*	*Significance*
-24 hours	-858 ± 73	-804 ± 111	0.54 (p=0.589)
0 hours	-739 ± 85	-533 ± 73	2.08 (p=0.039) *
4 hours	-944 ± 79	-479 ± 72	4.7 (p<0.001)***
8 hours	-740 ± 79	-486 ± 74	2.57 (p=0.011)*
12 hours	-802 ± 37	-301 ± 87	5.06 (p<0.001)***
24 hours	-895 ± 59	-454 ± 72	4.46 (p<0.001)***
48 hours	-985 ± 53	-542 ± 72	4.48 (p<0.001)***
72 hours	-961 ± 63	-425 ± 36	5.42 (p<0.001)***
96 hours	-1045 ± 74	-335 ± 63	7.18 (p<0.001)***

Net fluxes are measured as µmol/kg/h. Averages and standard error are shown for each population, as well as the significance of a two-way ANOVA accepting significance at $p<0.05$, where * ≤ 0.05; p ≤ 0.01 ** and p ≤ 0.001 ***

Chapter VI

Natural populations of chronically exposed metal-tolerant trout display minimal hepatic transcriptional responses to an acute metal exposure

Manuscript in preparation

AUTHORS: J.R. Paris[1], S. Shaw[2], R.A. King[1], D. Rowe[1], R.W. Wilson[1], N.R. Bury[3], E.M. Santos[1] & J.R. Stevens[1]

[1] Biosciences, College of Life and Environmental Sciences, University of Exeter, United Kingdom

[2] Centre for Genome Enabled Biology and Medicine, University of Aberdeen, Aberdeen, AB24 3FX, UK

[3] Faculty of Science, Health and Technology, University of Suffolk, Waterfront Building, Neptune Quay, Ipswich, IP4 1QJ, United Kingdom

J.R. Paris undertook the exposure experiments, conducted the laboratory work and the data analysis, and wrote the manuscript

S. Shaw assisted with data analysis, and contributed to writing of the manuscript

R.A. King assisted with the exposure experiments, undertook laboratory work and improved drafts of the manuscript

D. Rowe conceived the exposure experimental design and set up, maintained and assisted with the exposure experiments

R.W. Wilson conceived the exposure experiments

N.R. Bury conceived the exposure experiments

E.M. Santos conceived the exposure experiments, assisted with data analysis and improved drafts of the manuscripts

J.R. Stevens conceived the exposure experiments, provided support throughout and improved drafts of the manuscript

Natural populations of chronically exposed metal-tolerant trout display minimal hepatic transcriptional responses to an acute metal exposure

Manuscript in preparation

AUTHORS Paris JR[1], Shaw S[2], King RA[1], Rowe D[1], Stevens JR[1] & Santos EM[1]

[1] Biosciences, College of Life and Environmental Science, Geoffrey Pope Building, University of Exeter, Exeter, EX4 4QD, UK

[2] Centre for Genome Enabled Biology and Medicine, University of Aberdeen, Aberdeen, AB24 3FX, UK

ABSTRACT

The regulation of gene expression is considered a major contributor in determining the ability of fish to acclimate to polluted environments. Metal-tolerant brown trout (*Salmo trutta* L.) residing in a chronically contaminated river (River Hayle), are known to regulate the gene expression of key metal- and ion-binding genes in order to cope with the ecotoxicological effects of metal pollution in their environment. To date, the capacity of these fish to respond to exposure to a mixture of metals under laboratory conditions has never been investigated. We obtained fish from a metal-impacted river and a relatively unpolluted river and depurated them for 7 days in clean water, before exposing each group to a metal mixture (Cd, Cu, Ni and Zn). We quantified the transcriptional responses of each population using RNA-Seq in order to identify potential differences in the regulation of genes involved in the control of homeostasis of metals in each population. Importantly, there were no significant changes in gene transcription metal-tolerant fish following exposure to the metal mixture. In contrast, for naïve fish, the transcription of a suite of genes implicated in a generalised stress response were differentially regulated following exposure to the same metal mixture. In addition, immune system regulation appears to be constitutively over-expressed in the metal-tolerant fish, when compared to the control population. This dataset, therefore, provides compelling novel evidence for the tolerance of these fish to metal exposure.

KEYWORDS: RNA-seq, transcriptomics, brown trout, metals, metallothioneins, heat shock proteins, ecotoxicology

INTRODUCTION

There are over two million chemical compounds in current use, and approximately 1,000 new chemicals find their way into the environment each year (AFS Policy Statement #6). Among various toxic substances, metals are particularly severe in their action (Agrawal 2009). Aquatic habitats are particularly vulnerable to metal pollution, due to the high association of metals with sediment and organic material (Hart 1982; Miller 1997; Hochella *et al.* 2005; Jain *et al.* 2005). Aggregation of metals bound to sediments and organic material means that the dispersal distance of these pollutants is large and can contaminate the aquatic environment tens of kilometres away from the source of contamination (Salomons & Eagle 1990; Whiting *et al.* 2005). Impacted environments often suffer from reduced species diversity and significant degradation in ecological community structure, yet resilience to polluted environments has been shown to occur in several fish species (Hamilton *et al.* 2017).

Metals can differ widely in their interactions with biological systems; some are essential and are required for physiological functions and processes, yet become toxic at high concentrations. Non-essential metals, to date, have no recognised biological function, and are harmful to organisms even at low concentrations (Datta *et al.* 2009; Wang *et al.* 2014; Cobbina *et al.* 2015). In order to assess and quantify the toxic effects of metals, ecotoxicological studies have typically assessed responses to acute and chronic exposures of a single metal. In nature, however, metals rarely exist in isolation from another. The biological effects of one metal often influence the response of organisms to other metals or toxicants, with a large body of research illustrating the importance of biologically relevant toxic combinational effects (Paquin *et al.* 2000; Utgikar *et al.* 2004; Ibrahim & Haes 2005; Naddy *et al.* 2015).

The River Hayle, in southwest Britain, flows through an area characterised by prolific mining activity, at its height during the Industrial Revolution of the 18th

and 19th Centuries. As a consequence, the river still features high metal contamination from abandoned mining adits, in particular cadmium (Cd), copper (Cu), nickel (Ni) and zinc (Zi) (Minghetti *et al.* 2014). The river is classified as failing Environment Quality Standards (EQS), and its toxic metal concentrations are known to have decimated invertebrate and fish life across a ~2 km stretch of the river (Brown 1977; Durrant *et al.* 2011). Despite this, the river contains an apparently healthy and thriving population of brown trout (*Salmo trutta* L.), which endure chronic levels of metals throughout their lifetimes (Uren-Webster *et al.* 2013; Minghetti *et al.* 2014; Paris *et al.* 2015).

Fish are especially susceptible to the effects of metal pollution due to the intimate contact between gill tissue and the aquatic environment. Because of the range and admixture of metals that exist in a natural metal-impacted aquatic environment, fish are faced with balancing the uptake, transport and storage of essential metals against the prohibitive uptake, transport, detoxification and storage of non-essential metals (Wood *et al.* 2012). Metals most commonly enter the body via the gill and gut, through waterborne and dietary uptake routes, respectively (Playle 1998; Bury & Wood 1999; Evans *et al.* 1999; Bury *et al.* 2003). Once inside the fish, the metals are transported via the blood stream to various organs for processing, which are often metal specific (Farag *et al.* 1994; McGeer *et al.* 2000; Hollis *et al.* 2001).

Several studies have demonstrated that a large majority of metals accumulate in liver tissue (Agah *et al.* 2009; De Jonge *et al.* 2015; Jovičić *et al.* 2015) and that the liver is one of the most important organs involved in metal handling and processing for a broad range of metals (Langston *et al.* 2002; Hosseini *et al.* 2014; Rosabal *et al.* 2015). For the essential metals, Cu and Zn, the liver is fundamental in metal homeostasis and detoxification at toxic levels (Cousins 1985; Luza & Speisky 1986). Ni is a presumed essential metal, and less is known about its biological handling, yet activation of oxidative stress and antioxidant pathways have been observed in the liver (Zheng *et al.* 2014; Palermo *et al.* 2015). Cd is a non-essential metal, its toxicity primarily dealt with by the liver and kidney. Exposures show high Cd accumulations in the liver as well as induced hepatic oxidative damage (Yeşilbudak & Erdem 2014; Zheng *et*

al. 2016). The liver also plays a vital role in excretion of Cu and Cd via hepato-biliary routes, but less so for Ni and Zn (Wood *et al.* 2012).

Studying gene expression profiles has proven to be a highly effective method of quantifying genes and pathways involved in metal homeostasis and toxicity in fish (see recent review by Kumar & Denslow 2016). Previous research has shown that metallothioneins (MTs) and glutathione (GSH) are particularly important groups of proteins involved in metal scavenging and detoxification (Hogstrand & Haux 1991; Srikanth *et al.* 2013). Studies characterising gene expression in several tissues of brown trout from the River Hayle (Hayle trout) in comparison to trout from a clean river highlighted a suite of genes involved in metal-ion handling and homeostasis (Uren-Webster *et al.* 2013). This included one isoform of metallothionein (*mtb*), which was strongly upregulated in multiple tissues of Hayle trout, including the liver and the gills. Upregulation of *mtb* has thus been identified as an important metal-handling protein in conferring metal-tolerance in this population of trout, however, it is unknown how the behaviour of this, and other genes, change in Hayle trout following depuration in metal-free water, and how they are regulated in response to an acute metal-mixture exposure under laboratory conditions.

In order to better understand how tolerant fish respond to metal exposure, we conducted RNA-seq analysis on the livers of both naïve and tolerant brown trout after an acute 96-hour exposure to a mixture of both essential and non-essential metals (Cd, Cu, Ni and Zn). We explored differences in the transcriptional responses of tolerant and naïve fish and showed that metal-tolerant trout showed very limited changes in hepatic gene expression in responses to the exposure, yet naïve fish upregulated several metal-binding and stress-related genes following metal exposure.

MATERIALS AND METHODS

Fish sampling and laboratory maintenance

Sexually immature (0+ parr) wild brown trout were sampled from two rivers in Cornwall, the River Fowey (naïve population) and the River Hayle (tolerant population), in September 2015 by electrofishing, under Environment Agency

authorisation.

Fish were maintained under experimental conditions as specified in Chapter 5 of this thesis. Briefly, in order to avoid confounding effects caused by differences in water chemistry characteristics (pH, hardness, salinity, and alkalinity), fish were kept in a flow-through system of control water comprising salts matching natural-river chemistry. All fish were maintained for a depuration period of 7 days before being split into control and exposed cohorts, and being acutely exposed to a metal mixture comprising Cd (0.4 µg/L), Cu (3 µg/L), Ni (10 µg/L) and Zi (75 µg/L) for 96-hours. See Chapter 5 for further details.

At the end of the exposure period, liver tissue was sampled from naïve and tolerant fish from both exposed and control cohorts; tissues were snap-frozen in liquid nitrogen and stored at -80°C prior to analysis. A total of 24 individual fish were used across the four experimental groups: naïve-control (n=6); naïve-exposed (n=6); tolerant-control (n=6); tolerant-exposed (n=6). Although the fish were phenotypically identified as sexually immature, all fish were typed to determine their sex following the methods of Goodwin *et al.* (2016) and an equal number of males (n=3) and females (n=3) were sequenced from each experimental group to account for sex as a potential confounding factor during the bioinformatics analysis of the data.

Tissue extraction and library preparation

Liver tissue was weighed, homogenised in extraction buffer, and RNA was extracted using Qiagen RNA-seq extraction kits following the manufacturers' protocol, including an on-column DNAse digestion step. Total RNA quantification and quality assessment was conducted using a Bioanalyser (Agilent RNA 6000 Nano) and libraries were prepared following the TruSeq Stranded mRNA library preparation protocol (Illumina). Concentrations of cDNA libraries were quantified using a Tapestation (D1000 ScreenTape) and final library pool concentrations were determined by qPCR. Twenty-four libraries were pooled together and 125bp paired-end sequenced on two lanes of a standard run on the Illumina HiSeq 2500 at the University of Exeter Sequencing Service.

Bioinformatics quality control and processing

Raw reads were cleaned and processed using Trimmomatic (Bolger *et al.* 2014), using a quality threshold of 20. The first 15bp were trimmed from the 5' end of each read to prevent hexamer priming bias. Reads less than 30bp in length were discarded. Kmer distributions for each of the populations (naïve and tolerant) were computed using kmer_filter in Stacks (v1.42; Catchen *et al.* 2013) and plotted in Gnuplot v.5.0 (gnuplot.info). Based on the kmer frequency distribution plots, we removed rare kmers (--rare flag in kmer_filter) from the raw reads prior to normalisation. Reads were normalised to a maximum coverage threshold of 50 using *in silico* read normalisation in Trinity (Grabherr *et al* 2011).

Transcriptome assembly and annotation

De novo assembly of the data was performed using Trinity v2.2 (Grabherr *et al* 2011; Haas *et al.* 2013). Redundancy in the initial assembly was reduced using cd-hit-est (Li & Godzik 2006) and the final assembly was annotated using Trinotate v. 3.0 (https://trinotate.github.io/). Annotation was conducted following Trinotate guidelines with alignment to suggested databases (uniprot, transdecoder, pfam) using an e-value cut-off of $1e^{-5}$. We also generated custom blastx database annotations by alignment to zebrafish (*Danio rerio*) and Atlantic salmon (*Salmo salar*) peptide databases obtained from Ensembl (release 0.87), also using a p-value cut-off of $1e^{-15}$. A final transcriptome consisting only of functionally annotated transcripts was used downstream for read mapping and differential expression analyses. Assembly quality of the transcriptome was determined by (i) assessing the transcriptome statistics N50 and ExN50, (ii) calculating the proportion of reads that mapped back to the transcriptome using bowtie2 (Langmead & Salzberg 2012) and (iii) identifying the presence of Benchmarking Universal Single-Copy Orthologs (BUSCO; Simão *et al.* 2015).

Differential expression and GO term analysis

Transcript quantification was conducted by aligning transcripts to the transcriptome using bowtie v.1.1.1 (Langmead *et al.* 2009), reporting only the best hit (-k 1), and abundances were quantified using RSEM (Li & Dewey 2011). Differential gene expression between groups was determined using

edgeR (Robinson *et al.* 2010). We adopted a conservative approach to identifying differentially expressed genes, such that only genes which were differentially expressed by at least 2-fold at an FDR <0.05 (Benjamini-Hochberg correction) and present in at least 3 individuals (50% of each group) were included for downstream analysis. The data was summarised by using Principal Component Analysis (PCA) based on the list of all expressed genes using PtR in Trinity to visualise the major trends in the dataset.

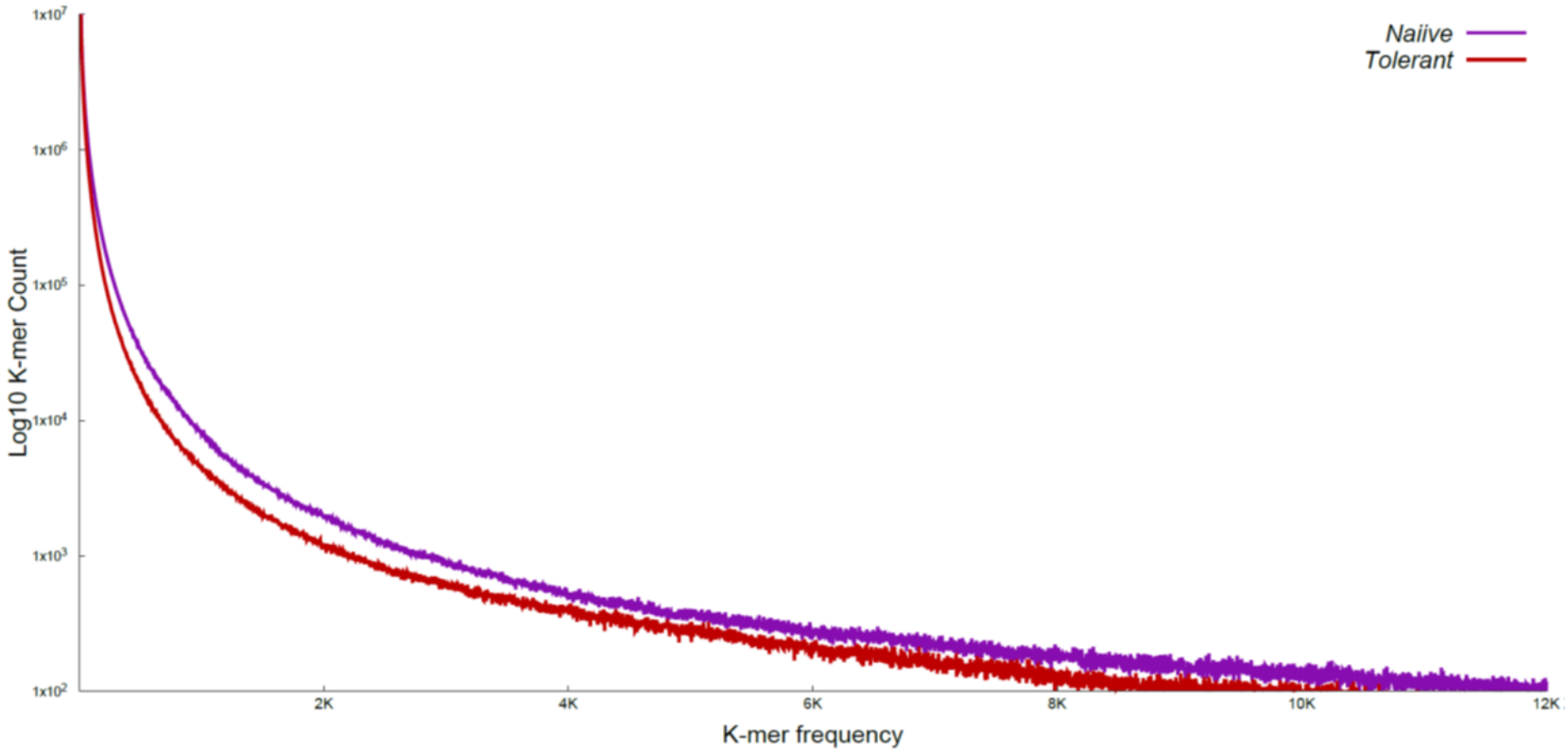
Differential expression between treatment groups was determined and analysis focused on three group comparisons: (i) naïve-control vs. naïve-exposed; (ii) tolerant-control vs. tolerant-exposed and (iii) naïve-exposed vs. tolerant-exposed. Heat maps based on the list of all differentially expressed genes were produced from the normalised count data using the heatmap.2 function of the gplots package (Warnes *et al.* 2009). For each comparison, over-representation of Gene ontology (GO) terms and KEGG (Kyoto Encyclopaedia of Genes and Genomes) pathways for the differentially expressed gene lists (considering up-regulated and down-regulated genes both separately and together) were identified using The Database for Annotation, Visualization and Integrated Discovery (DAVID v6.8; Huang *et al.* 2009ab), using all expressed transcripts as a background. Over-representation of Gene Ontology (GO) was calculated for Biological Process (BP), Cellular Component (CC) and Molecular Function (MF). GO terms or KEGG pathways were considered to be significantly enriched when Benjamini corrected p-values were <0.05. GO terms and their related gene annotations were plotted using GOplot v. 1.0.2 (Walter *et al.* 2015) in R v. 3.2.4 (R Development Core Team 2008).

RESULTS AND DISCUSSION

Quality control and transcriptome assembly

RNA-seq analysis of 24 liver samples generated over 442 million paired reads, averaging ~18 million reads per sample (SE ± 2,101,145 reads), of which ~278 million were retained after quality control, with an average of 11± 1,329,081 million reads per sample (Table S1). Kmer frequency distribution plots (Figure 1) showed that the reads of both the naïve and tolerant samples contained a

small number of rare kmers. These rare kmers were filtered out to reduce transcriptome redundancy introduced by polymorphism and sequencing error (Kelley *et al.* 2010; Brown *et al.* 2012).



.

Figure 1. Kmer distribution plots for the cleaned (not normalised) reads. Computed using kmer_filter for the two populations of fish: naïve (in purple) and tolerant (in red). Each read is reduced to a set of K-mers and appear in the distribution in proportion to the number of times a particular read was sequenced

The initial assembly comprised a total of ~254Mb, resulting in a final assembly of 357,444 transcripts corresponding to 280,186 putative genes. Redundancy in an assembly occurs when reads are assembled into multiple contigs, representing the same gene. Redundancy arises for a number of reasons, both biological (high polymorphism, indels and copy number variation), and assembly related (sequencing error, variation in read length). Such redundancy is undesirable because it compromises the ability to discriminate differential expression because several contigs derived from the same gene can be incorrectly identified as different differentially expressed genes (Ono *et al.* 2015). Removing redundancy in our initial assembly reduced the number of transcripts to 283,304. We used this assembly with reduced redundancy for annotation, and 42.5% of its transcripts were annotated to known protein domains and functions.

The final annotated transcriptome encompassed 118Mb and was comprised of 120,368 annotated transcripts (91,401 genes), with an average contig length of 980bp, an N50 of 1850bp and an ExN50 of 2270bp, (comprising 28,381 transcripts). Alignment of the reads to the transcriptome using bowtie2 showed that 94% mapped back as pairs of reads. A total of 3,023 core vertebrate BUSCO groups were searched, for which 2,259 groups were complete as single copy (63%) and multiple copies (12%), resulting in a BUSCO score of 75%. This score may appear low, but is not surprising given that the transcriptome was sequenced from a single tissue, and thus representation of all the core vertebrate gene groups would not be expected. This has been found in other single-tissue transcriptome experiments (Northcutt *et al.* 2016; Waiho *et al.* 2017). Taken together, these metrics indicate that the assembly was of good quality with over 90% of reads mapping back to the transcriptome assembly and 75% of core genes included in the final annotated transcriptome.

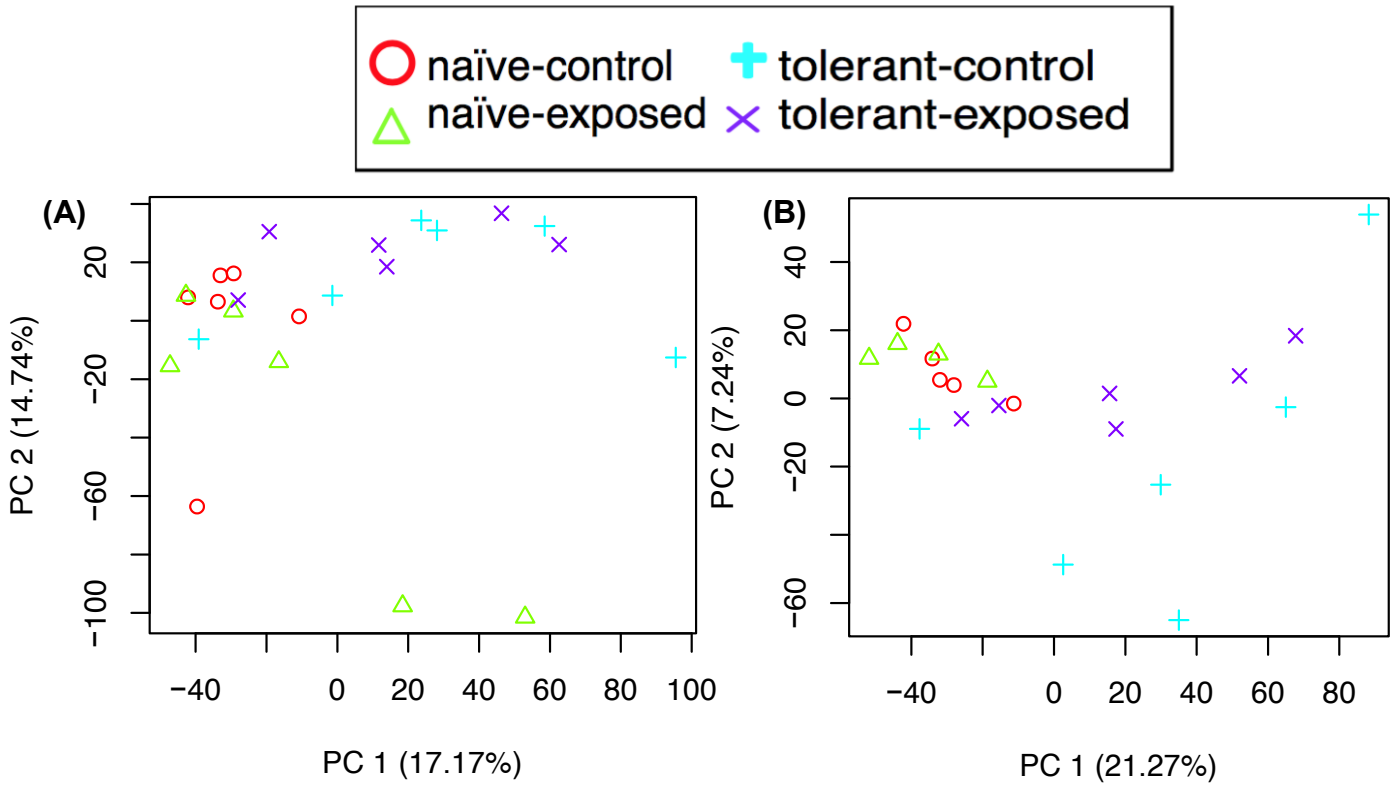


Figure 2. Principal Component Analyses (PCA) plots describing the log2 differentiation between biological replicates based on their gene expression profiles: (A) PCA plot drawn from all samples and replicates showing three outliers (naïve_control_06; naïve_exposed_04; naïve_exposed_06). (B) PCA plot drawn after removal of the three outlier samples.

Principal Components Analysis

Similarities between the biological replicates were visualised using PCA plots (Figure 2). Overall there was a separation between the naïve and tolerant fish, with the naïve fish showing distinct variation. PC1 accounted for 17.17% of the variance, differentiating between the naïve and tolerant fish. PC2 accounted for 14.74%; PC2 showed three-outlier sample replicates (Fig. 2A). Removal of the three outlier replicates (naïve-control 06; naïve-exposed 04; naïve-exposed 06) increased the differentiation between the naïve and tolerant samples to a PC1 value of 21.27% and showed that the naïve samples clustered more closely together, compared to the tolerant population, which showed more variance (Fig. 2B). Outliers such as these are removed to avoid confounding differential expression analyses.

Differential expression between control and exposed fish from a metal-tolerant and a metal-naïve population

The differential expression matrix contained a total of 900 differentially expressed genes (differentially expressed by more than 2 and FDR<0.05) across all pairwise comparisons. Figure 3 shows a heat map generated using normalised count data for all 900 differentially express genes and demonstrates the relationships between the expression of the two populations under control and exposed conditions. Expression was dominated by a large proportion of genes with modest changes in expression at this threshold (white; Fig. 3), and a very small number of genes with significantly strong changes in expression (red: under-expressed and blue: over-expressed; Figure 3). Notably, a large number of genes were observed to be over-expressed in the naïve population compared to the tolerant population.

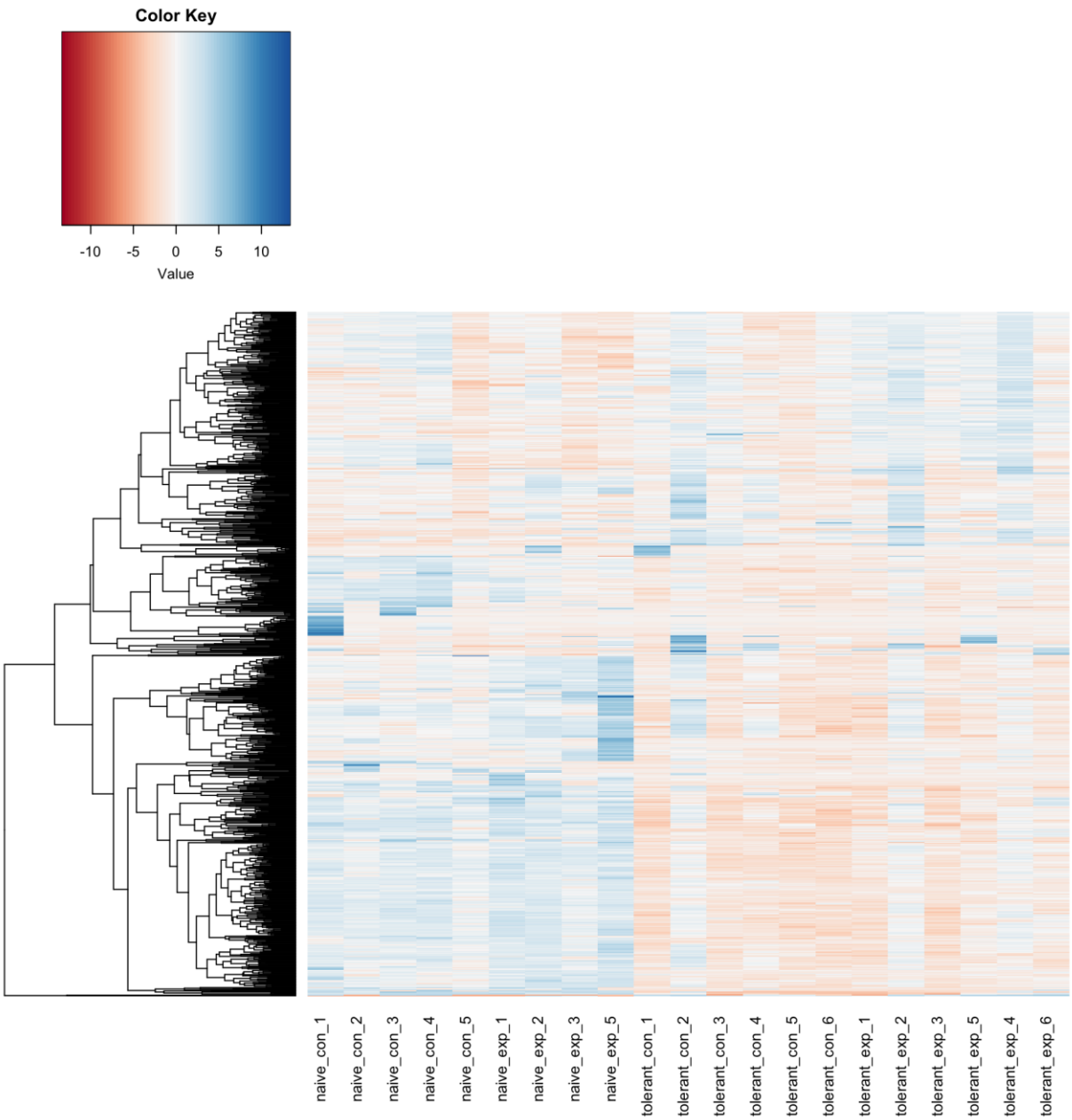


Figure 3. Heatmap of all differential expressed genes between treatment groups (≥2-fold change and FDR <0.05): naïve_control; naïve_exposed; tolerant_control; tolerant_exposed. White represents no DE expression, red represents under-expressed genes and blue represents over-expressed genes.

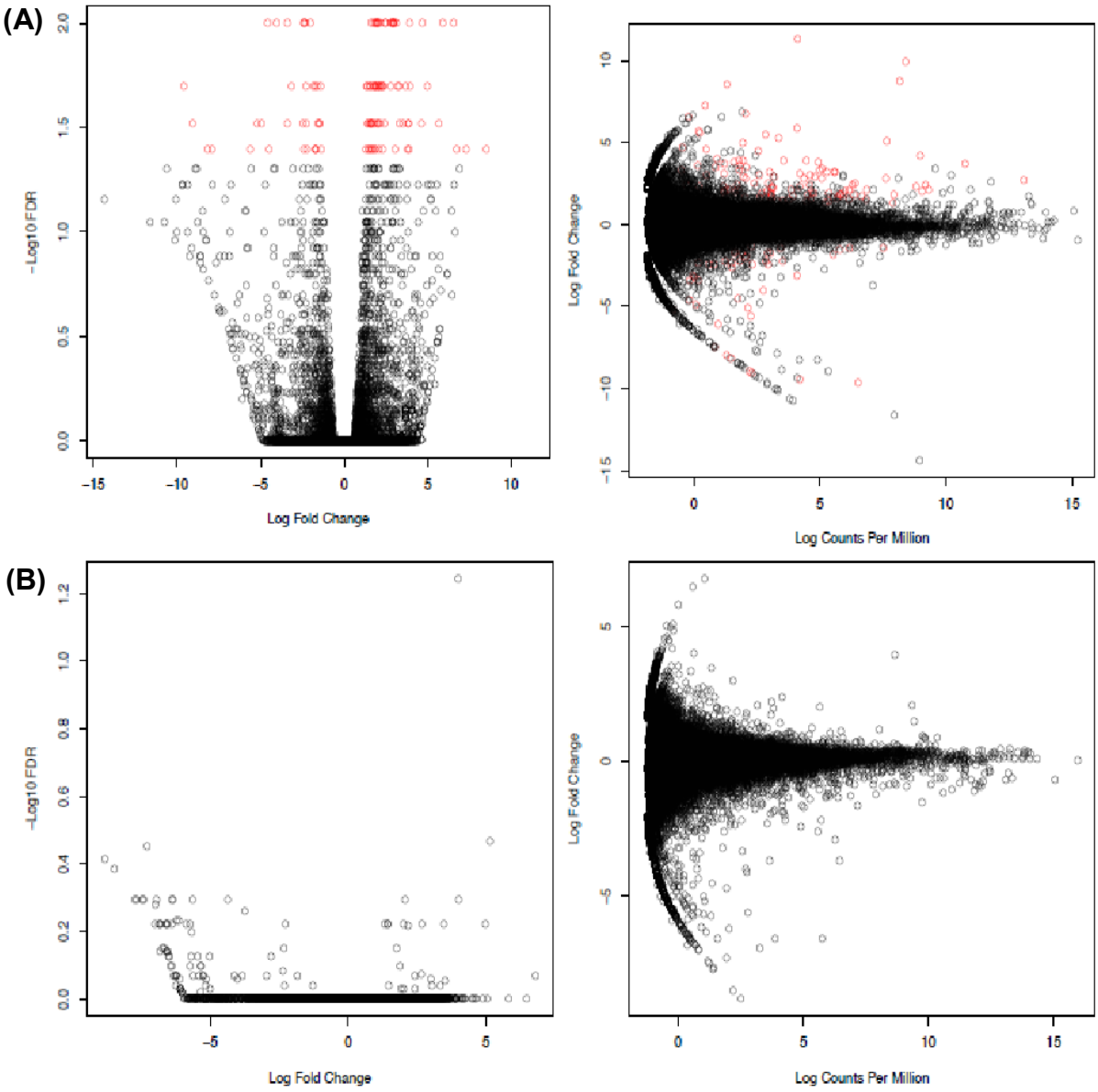


Figure 4. Plots of the differentially expressed transcripts between the following two comparisons: (A) naïve-control & naïve-exposed; and (B) tolerant-control & tolerant-exposed. Volcano plots represent the log fold change in gene expression against -log10 FDR and MA plots represent the log counts per million (M) against the mean log fold change (A) in expression. Significant differentially expressed genes (FDR<0.05) are represented in red.

A total of 112 genes were differentially expressed between naïve-control and naïve-exposed fish (Figure 4A); the majority of these, 78 genes (53 with gene annotations) were upregulated in the exposed fish (Table S2), with only 34 downregulated (23 with gene annotations) (Table S3). Strikingly, no genes were differentially expressed between tolerant-control and tolerant-exposed fish (Figure 4B), suggesting a greater tolerance of these fish to the metal exposure.

The gene with the largest over-expression (9.96-fold) in naïve fish as a result of the exposure to the metal mixture was leukocyte cell derived chemotaxin 2 (*lect2*), for which high concentrations have been associated with tumour suppression of hepatocellular cancer (Okabe *et al.* 2014; Chen *et al.* 2014). *lect2* has also been implicated as a generalised immune response to chemical toxicity in fish (Krasnov *et al.* 2007; Mori *et al.* 2007; Osuna-Jiménez *et al.* 2009). Other strongly upregulated genes included the transposable element Tcb2 transposase (*tbc2*; 8.76-fold), Tar1p (*tar1*; 6.89-fold) and ubiquitin C-terminal hydrolase (*ubp5*; 6.71-fold). A large number of the upregulated genes have functions related to metal-ion binding, in particular calcium- and magnesium- binding and these include *lect2; ubp5;* matrix metallopeptidase 9 (*mmp9);* phosphodiesterase 9A (*pde9a*)*;* myosin light chain kinase (*mylk*)*;* copine IV (*cpne4*)*;* intelectin (*itln*)*;* EPS15 homology domain (*ehd1*)*;* calreticulin (*calr*)*;* trypsinogen-like protein 3 (*trp3*) and ATPase sarcoplasmic/endoplasmic reticulum Ca^{2+} transporting 2 (*atp2a2*). As an example, *pde9a* (3.8-fold upregulated) is a high affinity phosphodiesterase enzyme (PDE), which contains a cofactor-binding site for zinc and magnesium (Liu *et al.* 2008); zinc is also known to modify and regulate PDE activity (Beyersmann & Haase 2001). *cpne4* (3.54-fold upregulated) is a Ca^{2+}-dependent phospholipid binding protein, which regulates ion transport at the cell membrane in response to Ca^{2+} fluxes (Jambunathan *et al.* 2001). *mylk* (3.65 fold upregulated) was upregulated in *O. mykiss* in response to handling stress (Krasnov *et al.* 2005), and *mmp9* (3.84-fold upregulated) is involved in metal-binding of both zinc and calcium (Tallant *et al.*2010), in addition to playing a key role in immune responses (Chadzinska *et al.* 2008).

Together, the over-expression of a wide range of genes involved in metal

binding suggests that this process is of particular importance for the response of naïve fish to metal exposure. The fact that similar up-regulation of these genes were absent in the metal-tolerant fish population suggests that other mechanisms are important in the tolerant population for conferring long-term resistance to metal toxicity or, alternatively, that the concentrations of metals used in this study were not sufficient to cause this biological response in metal-tolerant fish.

Naïve-exposed fish show enrichment of GO terms associated with metal stress

Exploring the over-represented GO terms within the list of upregulated genes following metal exposure in the naïve population revealed further insights into the biological mechanisms impacted by metal exposure. GO enrichment analysis of the upregulated genes identified 24 GO Biological Process (BP); 2 GO Molecular Function (MF) terms and 12 GO Cellular Component (CC) enriched terms (Table S4). The most enriched GO BP terms included: regulation of *release of cytochrome c from mitochondria* (49-fold), *ER-nucleus signalling pathway* (32-fold) and *ER-associated ubiquitin-dependent protein catabolic process* (21-fold). These over-represented GO terms represent processes typical of oxidative stress response (Craig *et al.* 2007; Krumschnabel *et al.* 2005), which is a commonly observed process associated with metal exposure (Sanchez *et al.* 2005; Valavanidis *et al.* 2006; Santos *et al.* 2010). On the other hand, GO analysis of the down-regulated genes did not identify any over-represented GO terms, potentially as a result of the gene list being relatively short or of many of its genes lacking in functional annotations.

Some of the key genes with several over-represented GO terms included *calr*, *mmp9*; *pla2g6* (calcium-independent phospholipase); the heat shock proteins *hsp90b1* (heat shock protein 90 beta family member 1) *hsp70* (heat shock 70kDa protein 5; *hspa5*) and their associated co-chaperones: *dnajc3* (DNAJ heat shock protein family member C3) and *dnajb11* (DNAJ heat shock protein family member B11). In particular, heat shock proteins (HSPs) are a family of proteins involved in the functioning of cells both under normal conditions and under stress (Parsell & Lindquist 1993) by regulating the folding and unfolding of polypeptides (Craig *et al.* 1993). HSPs are known to be upregulated as part

of general stress response (Iwama *et al.* 1998; Basu *et al.* 2002) and are frequently the first indicators in the response to metal toxicity (Bauman *et al.* 1993; Hansen *et al*. 2007; Jing *et al.* 2013). For example, *hsp70* is a molecular chaperone that forms an important part of the cellular response to oxidative stress (Padmini & Usha Rani 2008; Padmini & Vijaya Geetha 2009; Zeng *et al.* 2014; Figure 5). A number of metals are known to increase production of *hsp70*, (Boone & Vijayan 2002; Hallare *et al.* 2005; Lu *et al.* 2012; De Boeck *et al.* 2003). Constitutive forms of *hsp70* are present in unstressed cells, whereas the inducible form is synthesized by fish in response to stressors (Grosell 2012) with HSPs representing an important component of the metal-sensitive fraction (MSF) in fish. In natural populations of yellow perch (*Perca flavescens*) occupying metal-contaminated lakes affected by mining activity, it was shown that major proportions of hepatic Cd and Cu were associated with heat-stable cytosolic peptides and proteins fraction (HSP; Giguère *et al.* 2006) and *hsp70* gene expression was positively correlated with Cu accumulation in hepatic tissue (Pierron *et al.* 2009). This suggests that in metal-naïve fish HSPs are upregulated in response to the metal stress, whereas metal-tolerant fish may have constitutive forms of HSPs in their cells in order to protect themselves from chronic toxicity in their native River Hayle environment.

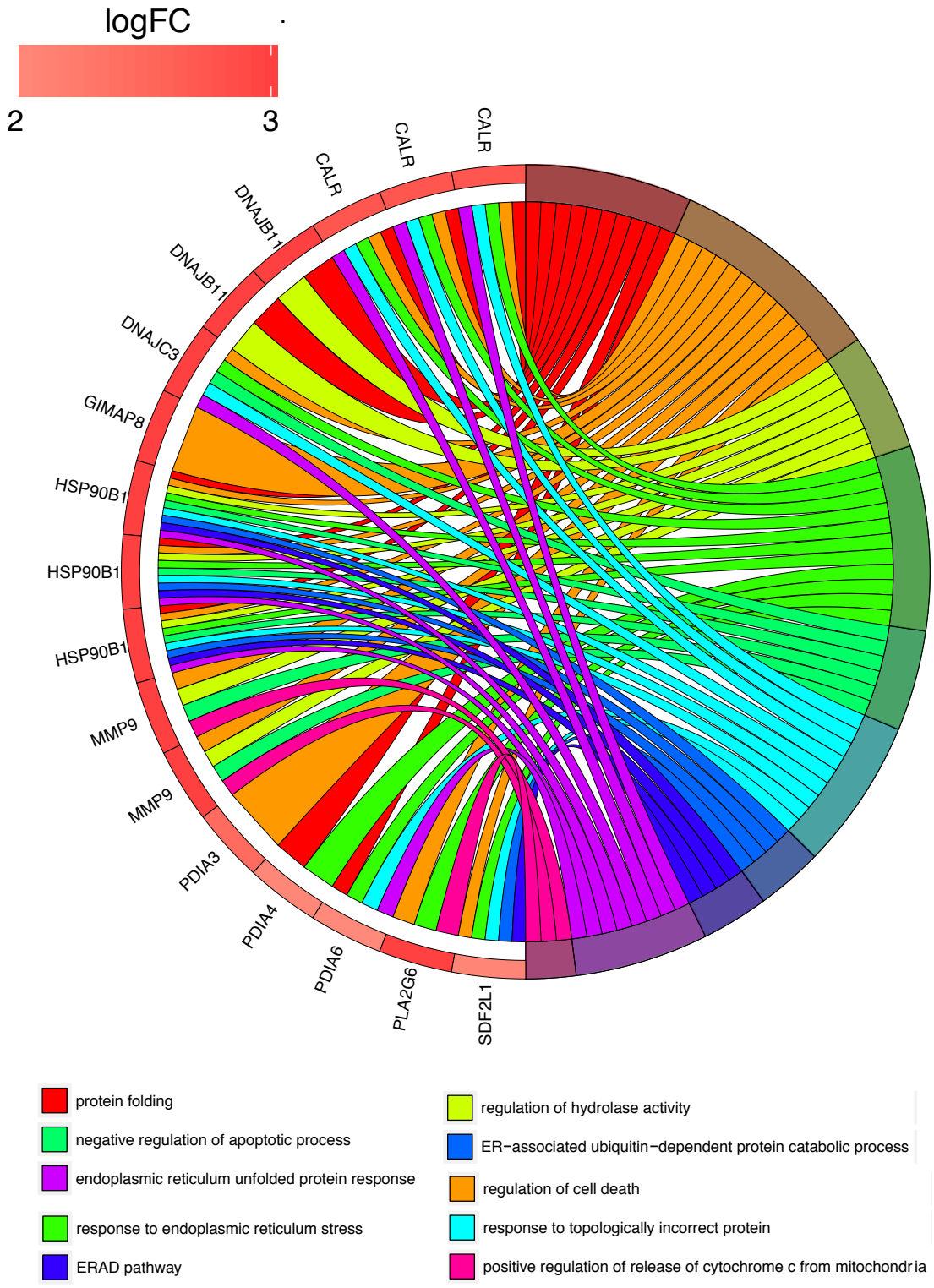
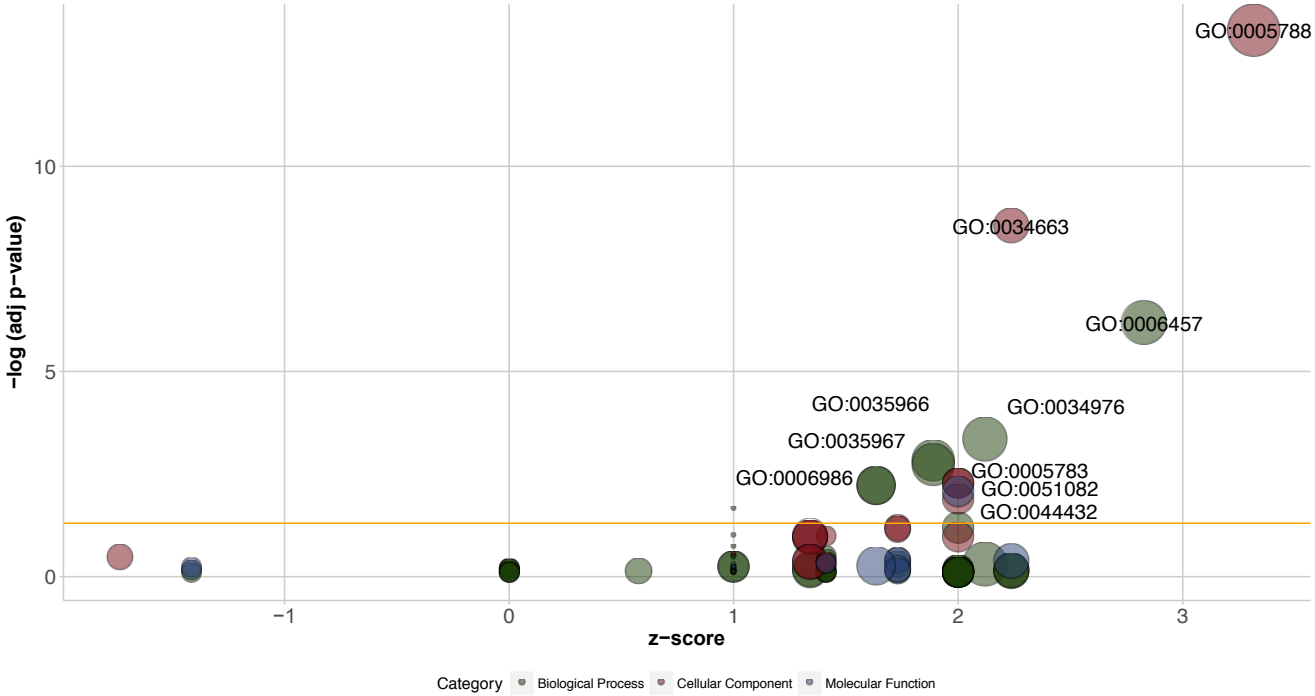


Figure 5. GO chord plot of the most significantly enriched genes in the GO Biological Process (BP) analysis of naïve-exposed fish.

GO BP enrichment among all differentially expressed genes in naïve fish revealed a number of enriched processes including *protein folding*, *regulation of cell death* and *response to endoplasmic reticulum stress* (Figure 6). These terms are linked to a general cell stress response to metal toxicity. For example, there is general consensus that proteins are key targets of metals, by (i) binding to free thiol groups or other functional groups within the protein, (ii) displacing essential metal ions in metalloproteins or (iii) catalysing oxidation of amino acid side chains (see review in Tamás *et al.* 2014).

Importantly, molecular chaperones such as HSPs play a significant role in recognising and binding to nascent polypeptide chains and partially folded intermediates of proteins, preventing their aggregation and misfolding (Fink 1999). *Protein folding* was also the most enriched GO BP term amongst the over-expressed genes identified in Hayle trout compared to a control population, in a previous study (Uren-Webster *et al.* 2013), as well as in other fish species exposed to metals (Mehinto *et al.* 2014; Hussainzada *et al.* 2014). Other enriched GO terms in the list of differentially expressed genes between naïve-control and naïve-exposed fish suggest a cascade of events leading to cell death: the biological response of *protein folding* triggers *endoplasmic reticulum stress*, and significant damage and prolonged stress is known to eventually trigger the process of *regulated cell death* (Kim *et al.* 2006; Fribley *et al.* 2009).



ID	Description
GO:0006457	protein folding
GO:0034976	response to endoplasmic reticulum stress
GO:0035966	response to topologically incorrect protein
GO:0035967	cellular response to topologically incorrect protein
GO:0006986	response to unfolded protein
GO:0030968	endoplasmic reticulum unfolded protein response
GO:0034620	cellular response to unfolded protein
GO:0031247	actin rod assembly
GO:0005788	endoplasmic reticulum lumen
GO:0034663	endoplasmic reticulum chaperone complex
GO:0005783	endoplasmic reticulum
GO:0044432	endoplasmic reticulum part
GO:0012505	endomembrane system
GO:0005789	endoplasmic reticulum membrane
GO:0042175	nuclear outer membrane−endoplasmic reticulum membrane network
GO:0048770	pigment granule
GO:0042470	melanosome
GO:0030496	midbody
GO:0051082	unfolded protein binding

Figure 6. GO Bubble plot of significantly enriched GO terms: Biological Process (BP; green); Cellular Component (CC; pink); Molecular Function (MF; blue). z-score is the mean expression of the differentially expressed genes in naïve-exposed fish. Size of the bubble represents the fold enrichment of those GO terms. GO term IDs represented in the bubble plot and present in the table represent GO terms with an adjusted p-value < 0.05.

Protein processing in the endoplasmic reticulum pathway in naïve-exposed fish

KEGG pathways over-represented among upregulated genes between naïve-control and naïve-exposed fish included *protein processing in endoplasmic reticulum* (*hsa04141*), which was over-represented by 27-fold. This pathway involves the biological processes of the endoplasmic reticulum (ER). Misfolded proteins are retained within the ER in association with chaperones, and if the misfolding is terminal, such proteins are degraded through ER-association degradation (ERAD). Genes overlapping with this pathway included *hsp90b1, dnajc3*, *calr*, protein disulphide isomerase family A members 4 and 6 (*pdia4* & *pdia6*). This pathway corroborates what was found in the enriched GO term analysis (Figure 6), suggesting metal exposure has caused oxidative stress associated with misfolded proteins and endoplasmic reticulum stress, resulting in the regulation of cell death.

Protein disulphide isomerases (*pdia6* & *pdia4*) are endoplasmic reticulum proteins which catalyse protein folding (Wilkinson & Gilbert 2004). Misfolded proteins are contained within the endoplasmic reticulum in complexes with molecular chaperones (Rao & Bredesen 2004), represented here by *calr* (a $Ca2^{+}$- binding chaperone that promotes protein folding) and HSPs (*hsp90b1*) and their associated co-chaperones (*dnajc3*). Proteins that are terminally misfolded are bound and are directed toward degradation through ER-associated degradation (ERAD; Vembar & Brodsky 2008). The ERAD pathway was also significantly enriched in the GO BP Terms (18% enrichment) and ubiquitin protein (*ubp5*) was also upregulated (6.7-fold). Accumulation of misfolded proteins in the ER causes ER stress and activates the unfolded protein response (UPR; Ron & Walter 2007). In certain severe situations, however, the protective mechanisms activated by the UPR are not sufficient to restore normal ER function, thereby resulting in the *regulation of cell death* via apoptosis. ERAD processes and endoplasmic reticulum stress were significantly enriched in stress responses of grouper species (*Epinephelus* spp.) to viral pathogens (Lu *et al.* 2012) and in the goby *Synechogobius hasta,* in response to a Cu exposure *(Song et al. 2016)*. RNA-seq analysis of thermal tolerance in catfish indicated the importance of genes associated with protein

folding and degradation in maintaining heat-tolerance (Liu *et al.* 2013). In sum, these data together suggest that metal exposure in naïve fish particularly targets protein stability, a process that does not appear to be affected in metal-tolerant fish.

Lack of induction of metallothioneins and the frontloading hypothesis

Interestingly, *mtb*, which was strongly upregulated in tolerant fish compared to naïve fish in an in-river gene expression analysis associated with metal tolerance (Uren-Webster *et al.* 2013), was not differentially expressed in any of the pairwise comparisons in this study (naïve-control vs naïve-exposed; tolerant-control vs tolerant-exposed; naïve-exposed vs tolerant-exposed). The transcript TRINITY_DN137158 was annotated as *mtb*, and, although expressed in the hepatic tissue of all groups, did not show significant differences between groups of naïve fish (logFC 0.82; logCPM 8.2; FDR 0.78) or tolerant fish (logFC 0.62; logCPM 7.8; FDR 1) in response to metal exposure. In this study, the fish had been depurated in clean water for 7 days, and the metal exposure over the 4-day period involved metal concentration levels significantly lower than those present in the native river (River Hayle) of tolerant trout: river Cd 1.4 µg/L vs. laboratory Cd 0.4ug/L; river Cu 43 µg/L vs. laboratory 3 µg/L; river Ni 27 µg/L vs laboratory 10 µg/L; river Zn 631 µg/L vs laboratory Zn 75 µg/L). This suggested that the short-term metal exposure encountered in the laboratory was not sufficiently high or sufficiently prolonged to elicit a metallothionein response in the liver of these fish.

The lack of upregulation of *mtb* highlights two interesting findings; firstly, that metallothioneins (MTs) are not necessarily upregulated when fish are exposed to a short-term exposure (96 hours) to low concentrations of a mixture of essential and non-essential metals. A similar response has been observed in zebrafish (*Danio rerio*) exposed to Cu concentrations of 8 µg/L and 15 µg/L for 48 hours (Craig *et al.* 2007). Misra *et al.* (1989) showed that in chinook salmon (*Oncorhynchus tshawytscha*) embryonic fish cell lines exposed to 24-72 hours of 20 µM $CdCl_2$ and 6-14 hours of 150 µM $ZnCl_2$, did not induce MT expression, yet heat shock proteins (HSPs) showed significant induction. It has been hypothesised that MTs are specifically induced in response to toxic metal

concentrations, their induction being both concentration- and time-dependent (Roesijadi 1994; De Boeck *et al.* 2003). On the other hand, the synthesis of HSPs are recognised as a general stress response (Misra *et al.* 1989; Boone & Vijayan 2002) and the upregulation of MTs and HSPs are often not correlated, with their regulation dependent on the exposure and metal-type (Bauman *et al.* 1993; Haap *et al.* 2016). Therefore, the lack of MT hepatic expression observed here may be the result of the low concentrations used, in addition to the short period over which the fish were exposed.

Secondly, these data suggest that during the weeklong depuration period, followed by acute exposure to significantly lower metal concentrations than the tolerant fish experience in-river, metal-tolerant trout have likely reduced the expression of *mtb* in liver*,* although metallothionein proteins may still be present in high quantities. Expression of this gene is considered to be vital in their natural environment, but it appears that metallothionein is no longer over-expressed following a period of depuration in the laboratory – i.e. expression is not being 'frontloaded' (Schoville *et al.* 2012). Such 'frontloading' has been proposed as an adaptive strategy under a variety of stressful environmental conditions (Schoville *et al.* 2012; Barshis *et al.* 2013). However, given the energy expenditure involved in over-expressing genes unnecessarily, this is a potentially costly strategy for long-term resilience (Wagner 2005; Sokolova *et al.* 2012). Indeed, the liver metal accumulation patterns of the tolerant fish used in this experiment showed that Cd, Cu and Ni concentrations significantly depurated from hepatic tissue after a week of the fish being maintained in clean freshwater, although concentrations of Cu did significantly increase in hepatic tissue at 96-hours exposed (Paris *et al.* unpublished/Chapter 5 of this thesis). However, metallothionein induction alone may not be necessary for these fish to cope with concentrations of metals significantly lower than those experienced in-river.

Alternatively, metal handling may be processed in other organs, meaning *mtb* could be over-expressed in tissues not examined here. The gills and kidney, for example, play an important role in metal homeostasis, and *mtb* has been shown to be strongly expressed in these organs in response to metals (Hermesz *et al.*

2001; Alvarado *et al.* 2006; Huang *et al.* 2007). In particular, short-term exposures appear to elicit a significant metallothionein response in gill tissue, whereas such changes may not be observed in the liver within the same time scale (Heerden *et al.* 2004; Laing *et al., unpublished*; Fitzgerald *et al., unpublished)*. Furthermore, we did not identify the other isoform of metallothionein (*mta*) in the transcriptome, suggesting it was either not expressed in the liver, or alternatively, that the gene was not assembled appropriately, probably due to low read counts representing this gene.

In order to further explore the mechanisms by which tolerant fish cope with the metal exposure, we also examined the differentially expressed transcripts between naïve-exposed and tolerant-exposed cohorts (Table S5). The most over-expressed transcript in tolerant-exposed fish was not annotated using the e-value cut-off used for the annotation of the transcriptome ($1e^{-5}$), but blastn of the transcript revealed high sequence identity (86%) to the Atlantic salmon (*Salmo salar*) IgH locus A (immunoglobin heavy chain locus isolocus A; Genebank: GU129139). Of note, this transcript was not identified as upregulated in the naïve fish. Immunoglobulin is one of the most important molecules in adaptive immunity (Hikima *et al.* 2011). The expression of the IgH locus has been implemented as playing a central role in a generalised immune response in salmonids (Tadiso *et al.* 2011; Xu *et al.* 2015), whilst innate immunology is purported to play a vital role in immunomodulation of metal toxicity (Zelikoff 1993; Jokinen *et al. 1995;* Sanchez-Dardon *et al.* 1999; Fatima *et al.* 2000). Furthermore, other significantly over-expressed transcripts related to immune-system function and responses were found including *mr1* (MHC class I-related gene protein; 5.8-fold); *bcl6* (B-cell lymphoma 6; 2.8-fold); *themis* (thymocyte selection associated molecule; 2.7-fold), suggesting that a constitutively expressed general immune response may be important for metal-tolerance in Hayle trout.

A suite of metal-binding genes were over-expressed in tolerant, compared to naïve fish including LINE-1 retrotransposable element ORF2 protein (*ORF2p*; 6.6-fold); phosphoribosyl-glycinamide synthetase (*gart*; 5.5-fold); zinc finger protein 569 (*znf569* (4.37-fold); ras association domain-containing protein

(*rassf1*; 3.2-fold); betaine-homocysteine methyltransferase (*bhmt*; 2.8-fold), and seven transcripts were annotated as serum albumin (three as serum albumin 1, *alb1* and four as serum albumin 2, *alb2*), which all showed significant over-expression (by 2.28 – 3.98-fold). Serum albumins are large multi-domain proteins, which are the primary protein constituents of the circulatory system, and have several physiological functions (He & Carter 1992). In particular, the proteins serve an important role in the transport of endogenous and exogenous compounds (Kragh-Hansen 1981), including several metals (Masuoka *et al.* 1993; Bal *et al.* 1998; Sweet & Zelikof 2010). In fish, increases in serum albumins have been used as a biomarker for various pollutants (Gopal *et al.* 1997; Fırat & Kargın 2010; Javed & Usmani 2015). Lastly, the cytochrome P450 gene (*cyp3a27*) was also upregulated (2.3-fold). This gene has important functions in regulating the metabolism of xenobiotic and environmental toxicants (see Bucheli & Fent 1995). Therefore, a general alteration from expression of HSPs to immune response genes could demonstrate the shift of the biological handling of metals from stress response in metal-naïve trout to acclimation in metal-tolerant trout.

In summary, we have found that acute exposure to a mixture of metals does not affect hepatic gene expression in trout from a metal contaminated river, whereas metal-naïve fish show significant upregulation of genes involved in metal and ion-binding, oxidative stress and a generalised stress response. Biological analysis of the differentially expressed genes suggests acute exposure affects the endoplasmic reticulum stress pathway in naïve fish but not in metal-tolerant fish. In this short-term exposure to relatively low concentrations of metals, metallothionein expression was not induced in either of the populations, despite playing a significant role in metal-tolerant fish in their natural habitat. Instead, it appears that genes involved in immune response may participate in the long-term resilience to metals in these metal-tolerant fish populations.

ACKNOWLEDGEMENTS

This work was funded by the University of Exeter, the Environment Agency and the Westcountry Rivers Trust. We thank the Fisheries Society of the British Isles for a Small Research Grant, which funded the RNA-seq library preparation and sequencing costs. Thanks to Dr Konrad Paszkiewicz, Dr Karen Moore and Audrey Farbos from the Exeter Sequencing Service for RNA-seq library preparation. Thanks to Lab 201 and the Aquatic Resource Centre for assisting with the experimental exposure, especially Dr Gregory Paull, Steve Cooper, and John Dowdle, and also to Dr Anke Lange, Dr Jennifer Fitzgerald and Lauren Laing for assistance with sampling and experimental design. Thanks also to Dr Ronny van Aerle, Dr Bas Verbruggen and Dr Tamsyn Uren-Webster for suggestions and guidance in the RNA-seq data analysis. Thanks also to Matt Healey, Giles Ricard and Dr Bruce Stockley from the Westcountry Rivers Trust and also to Jan and Phil Shears for organising the electrofishing and assisting with fieldwork for the fish collection.

REFERENCES

Agah, H., Leermakers, M., Elskens, M., Fatemi, S. M. R., & Baeyens, W. (2009). Accumulation of trace metals in the muscle and liver tissues of five fish species from the Persian Gulf. *Environmental Monitoring and Assessment*, **157**(1-4), 499–514.

Agrawal, S.K. (2009). Heavy Metal Pollution. APH Publishing Corporation: New Dehli

Alvarado, N. E., Quesada, I., Hylland, K., Marigómez, I., & Soto, M. (2006). Quantitative changes in metallothionein expression in target cell-types in the gills of turbot (*Scophthalmus maximus*) exposed to Cd, Cu, Zn and after a depuration treatment. *Aquatic Toxicology*, **77**(1), 64–77.

American Fisheries Society (AFS). Effects of toxic substances in surface waters. AFS Policy Statement #6. Accessed from fisheries.org/policy-media/policy-statements/afs-policy-statement-6/

Barshis, D. J., Ladner, J. T., Oliver, T. A., Seneca, F. O., Traylor-Knowles, N., & Palumbi, S. R. (2013). Genomic basis for coral resilience to climate change. *Proceedings of the National Academy of Sciences of the United States of America*, **110**(4), 1387–92.

Basu, N., Todgham, A. E., Ackerman, P. A., Bibeau, M. R., Nakano, K., Schulte, P. M., & Iwama, G. K. (2002). Heat shock protein genes and their functional significance in fish. *Gene*, **295**(2), 173–183.

Bal, W., Christodoulou, J., Sadler, P. J., & Tucker, A. (1998). Multi-metal binding site of serum albumin. *Journal of Inorganic Biochemistry*, **70**(1),

33–39.

Bauman, J. W., Liu, J., & Klaassen, C. D. (1993). Production of metallothionein and heat-shock proteins in response to metals. *Toxicological Sciences*, **21**(1), 15–22.

Beyersmann, D., & Haase, H. (2001). Functions of zinc in signaling, proliferation and differentiation of mammalian cells. *BioMetals*, **14**(3/4), 331–341.

Bode, W., & Maskos, K. (2003). Structural basis of the matrix metalloproteinases and their physiological inhibitors, the tissue inhibitors of metalloproteinases. *Biological Chemistry*, **384**(6), 863–72.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**(15), 2114–2120.

Boone, A. N., & Vijayan, M. M. (2002). Constitutive heat shock protein 70 (HSC70) expression in rainbow trout hepatocytes: effect of heat shock and heavy metal exposure. *Comparative Biochemistry and Physiology. Toxicology & Pharmacology : CBP*, **132**(2), 223–33.

Brown, B. (1977). Effects of mine drainage on the River Hayle, Cornwall a) factors affecting concentrations of copper, zinc and iron in water, sediments and dominant invertebrate fauna. *Hydrobiologia*, **52**, 2–3.

Brown, C. T., Howe, A., Zhang, Q., Pyrkosz, A. B., & Brom, T. H. (2012). A Reference-Free Algorithm for Computational Normalization of Shotgun Sequencing Data. Retrieved from http://arxiv.org/abs/1203.4802.

Bucheli, T. D., & Fent, K. (1995). Induction of cytochrome P450 as a biomarker for environmental contamination in aquatic ecosystems. *Critical Reviews in Environmental Science and Technology*, **25**(3), 201–268.

Bury, N. R., Walker, P. A., & Glover, C. N. (2003). Nutritive metal uptake in teleost fish. *Journal of Experimental Biology*, **206**(1).

Bury, N. R., & Wood, C. M. (1999). Mechanism of branchial apical silver uptake by rainbow trout is via the proton-coupled Na^+ channel. *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology*, **277**(5).

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, **22**(11), 3124-3140.

Chadzinska, M., Baginski, P., Kolaczkowska, E., Savelkoul, H. F. J., & Kemenade, B. M. L. V. (2008). Expression profiles of matrix metalloproteinase 9 in teleost fish provide evidence for its active role in initiation and resolution of inflammation. *Immunology*, **125**(4), 601–10.

Chen, C.K., Yang, C.Y., Hua, K.T., Ho, M.C., Johansson, G., Jeng, Y.M., Chen, C.N., Chen, M.W., Lee, W.J., Su, J.L., Lai, T.C., Chou, C.C., Ho, B.C., Chang, C.F., Lee P.H., Chang, K.J., Hsiao, M., Lin, M.T., & Kuo, M.L. (2014). Leukocyte cell-derived chemotaxin 2 antagonizes MET receptor

activation to suppress hepatocellular carcinoma vascular invasion by protein tyrosine phosphatase 1B recruitment. *Hepatology*, **59**(3), 974–985.

Cobbina, S. J., Chen, Y., Zhou, Z., Wu, X., Feng, W., Wang, W., Mao, G., Xu, H., Zhang, Z., Wu, Z., & Yang, L. (2015). Low concentration toxic metal mixture interactions: Effects on essential and non-essential metals in brain, liver, and kidneys of mice on sub-chronic exposure. Chemosphere, **132**, 79–86.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Wojciech S. M., Gaffney, D.J., Elo, L.L., Zhang, X., & Mortazavi, A. 2016. A survey of best practices for RNA-seq data analysis. *Genome Biology*, **17**(13).

Cousins, R. J. (1985). Absorption, transport, and hepatic metabolism of copper and zinc: special reference to metallothionein and ceruloplasmin. *Physiological Reviews*, **65**(2).

Craig, E. A., Gambill, B. D., & Nelson, R. J. (1993). Heat shock proteins: molecular chaperones of protein biogenesis. *Microbiological Reviews*, **57**(2), 402–14.

Craig, P. M., Wood, C. M., & McClelland, G. B. (2007). Oxidative stress response and gene expression with acute copper exposure in zebrafish (Danio rerio). *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology*, **293**(5).

Datta, S., Ghosh, D., Saha, D. R., Bhattacharaya, S., & Mazumder, S. (2009). Chronic exposure to low concentration of arsenic is immunotoxic to fish: Role of head kidney macrophages as biomarkers of arsenic toxicity to Clarias batrachus. Aquatic Toxicology, **92**(2), 86–94.

De Boeck, G., De Wachter, B., Vlaeminck, A., & Blust, R. (2003). Effect of cortisol treatment and/or sublethal copper exposure on copper uptake and heat shock protein levels in common carp, *Cyprinus carpio*. *Environmental Toxicology and Chemistry*, **22**(5), 1122–1126.

De Jonge, M., Belpaire, C., Van Thuyne, G., Breine, J., & Bervoets, L. (2015). Temporal distribution of accumulated metal mixtures in two feral fish species and the relation with condition metrics and community structure. *Environmental Pollution*, **197**, 43–54.

Durrant, C. J., Stevens, J. R., Hogstrand, C., & Bury, N. R. (2011). The effect of metal pollution on the population genetic structure of brown trout (*Salmo trutta L.)* residing in the River Hayle, Cornwall, UK. *Environmental Pollution,* **159**(12), 3595–603.

Evans, D. H., Piermarini, P. M., & Potts, W. T. W. (1999). Ionic transport in the fish gill epithelium. *Journal of Experimental Zoology*, **283**(7), 641–652.

Farag, A. M., Boese, C. J., Bergman, H. L., & Woodward, D. F. (1994). Physiological changes and tissue metal accumulation in rainbow trout

exposed to foodborne and waterborne metals. *Environmental Toxicology and Chemistry*, **13**(12), 2021–2029.

Fatima, M., Ahmad, I., Sayeed, I., Athar, M., & Raisuddin, S. (2000). Pollutant-induced over-activation of phagocytes is concomitantly associated with peroxidative damage in fish tissues. *Aquatic Toxicology*, **49**(4), 243–250.

Fink, A. L. (1999). Chaperone-mediated protein folding, **79**(2), 425–449.

Fırat, Ö., & Kargın, F. (2010). Individual and combined effects of heavy metals on serum biochemistry of Nile tilapia *Oreochromis niloticus*. *Archives of Environmental Contamination and Toxicology*, **58**(1), 151–157.

Fribley, A., Zhang, K., & Kaufman, R. J. (2009). Regulation of apoptosis by the unfolded protein response. In *Methods in molecular biology (Clifton, N.J.)* (Vol. 559, pp. 191–204).

Giguère, A., Campbell, P., Hare, L., & Couture, P. (2006). Sub-cellular partitioning of cadmium, copper, nickel and zinc in indigenous yellow perch (*Perca flavescens*) sampled along a polymetallic gradient. *Aquatic Toxicology*, **77**(2), 178–189.

Gopal, V., Parvathy, S., & Balasubramanian, P. R. (1997). Effect of heavy metals on the blood protein biochemistry of the fish *Cyprinus carpio* and its use as a bio-indicator of pollution stress. *Environmental Monitoring and Assessment*, **48**(2), 117–124.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdbury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindbald-Toh, K., Friedman, N., & Regev A. (2011). Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nature Biotechnology*, **29**(7), 644–652.

Grosell, M. 2012. Copper. In: Wood, C., Farrell, A.P., and Brauner C.J. eds., *Homeostasis and Toxicology of Essential Metals*. Oxford: Elsevier Academic Press, pp. 54-110.

Haap, T., Schwarz, S., & Köhler, H.-R. (2016). Metallothionein and Hsp70 trade-off against one another in *Daphnia magna* cross-tolerance to cadmium and heat stress. *Aquatic Toxicology*, **170**, 112–119.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J.,Couger, M.B., Eccles, D., Li, B., Lieber, M., Macmanes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., Leduc, R.D., Friedman, N., & Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, **8**(8), 1494–1512.

Hallare, A. V., Schirling, M., Luckenbach, T., Köhler, H.R., & Triebskorn, R. (2005). Combined effects of temperature and cadmium on developmental parameters and biomarker responses in zebrafish (*Danio rerio*) embryos.

Journal of Thermal Biology, **30**(1), 7–17.

Hamilton, P. B., Rolshausen, G., Uren Webster, T. M., & Tyler, C. R. (2016). Adaptive capabilities and fitness consequences associated with pollution exposure in fish. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **372**(1712).

Hansen, B. H., Garmo, O. A., Olsvik, P. A., & Andersen, R. A. (2007). Gill metal binding and stress gene transcription in brown trout (*Salmo trutta*) exposed to metal environments: the effect of pre-exposure in natural populations. *Environmental Toxicology and Chemistry*, **26**(5), 944–53.

Hart, B. T. (1982). Uptake of trace metals by sediments and suspended particulates: a review. *Hydrobiologia*, **91-92**(1), 299–313.

He, X. M., & Carter, D. C. (1992). Atomic structure and chemistry of human serum albumin. *Nature*, **358**(6383), 209–215.

Hermesz, E., Abrahám, M., & Nemcsók, J. (2001). Tissue-specific expression of two metallothionein genes in common carp during cadmium exposure and temperature shock. *Comparative Biochemistry and Physiology. Toxicology & Pharmacology : CBP*, **128**(3), 457–65.

Hikima, J., Jung, T.-S., & Aoki, T. (2011). Immunoglobulin genes and their transcriptional control in teleosts. *Developmental & Comparative Immunology*, **35**(9), 924–936.

Hochella, M. F., Moore, J. N., Putnis, C. V., Putnis, A., Kasama, T., & Eberl, D. D. (2005). Direct observation of heavy metal-mineral association from the Clark Fork River Superfund Complex: Implications for metal transport and bioavailability. *Geochimica et Cosmochimica Acta*, **69**(7), 1651–1663.

Hogstrand, C., & Haux, C. (1991). Binding and detoxification of heavy metals in lower vertebrates with reference to metallothionein. *Comparative Biochemistry and Physiology Part C: Comparative Pharmacology*, **100**(1-2), 137–141.

Hollis, L., Hogstrand, C., & Wood, C. M. (2001). Tissue-specific cadmium accumulation, metallothionein induction, and tissue zinc and copper levels during chronic sublethal cadmium exposure in juvenile rainbow trout. *Archives of Environmental Contamination and Toxicology*, **41**(4), 468–474.

Hosseini, M.-J., Shaki, F., Ghazi-Khansari, M., & Pourahmad, J. (2014). Toxicity of copper on isolated liver mitochondria: impairment at complexes I, II, and IV leads to increased ROS production. *Cell Biochemistry and Biophysics*, **70**(1), 367–381.

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*, **37**(1), 1–13.

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, **4**(1), 44–57.

Huang, Z.-Y., Zhang, Q., Chen, J., Zhuang, Z.-X., & Wang, X.-R. (2007). Bioaccumulation of metals and induction of metallothioneins in selected tissues of common carp (*Cyprinus carpio L.)* co-exposed to cadmium, mercury and lead. *Applied Organometallic Chemistry*, **21**(2), 101–107.

Hussainzada, N., Lewis, J. A., Baer, C. E., Ippolito, D. L., Jackson, D. A., & Stallings, J. D. (2014). Whole adult organism transcriptional profiling of acute metal exposures in male Zebrafish. *BMC Pharmacology and Toxicology*, **15**(1), 15.

Ibrahim, M., & Haes, H. El. (2005). Computational spectroscopic study of copper, cadmium, lead and zinc interactions in the environment. *International Journal of Environment and Pollution*, **23**(4), 417.

Iwama, G. K., Thomas, P. T., Forsyth, R. B., & Vijayan, M. M. (1998). Heat shock protein expression in fish. *Reviews in Fish Biology and Fisheries*, **8**(1), 35–56.

Jain, C. K., Singhal, D. C., & Sharma, M. K. (2005). Metal pollution assessment of sediment and water in the River Hindon, India. *Environmental Monitoring and Assessment*, **105**(1-3), 193–207.

Jambunathan, N., Siani, J. M., & McNellis, T. W. (2001). A humidity-sensitive *Arabidopsis copine* mutant exhibits precocious cell death and increased disease resistance. *The Plant Cell*, **13**(10), 2225–40.

Javed, M., & Usmani, N. (2015). Stress response of biomolecules (carbohydrate, protein and lipid profiles) in fish *Channa punctatus* inhabiting river polluted by Thermal Power Plant effluent. *Saudi Journal of Biological Sciences*, **22**(2), 237–242.

Jing, J., Liu, H., Chen, H., Hu, S., Xiao, K., & Ma, X. (2013). Acute effect of copper and cadmium exposure on the expression of heat shock protein 70 in the Cyprinidae fish *Tanichthys albonubes*. *Chemosphere*, **91**(8), 1113–1122.

Jokinen, E. I., Aaltonen, T. M., & Valtonen, E. T. (1995). Subchronic Effects of pulp and paper mill effluents on the immunoglobulin synthesis of roach, *Rutilus rutilus*. *Ecotoxicology and Environmental Safety*, **32(**3), 219–225.

Jovičić, K., Nikolić, D. M., Višnjić-Jeftić, Ž., Đikanović, V., Skorić, S., Stefanović, S. M. Lenhardt, M., Hegediš, A., Krpo-Ćetković, J., & Jarić, I. (2015). Mapping differential elemental accumulation in fish tissues: assessment of metal and trace element concentrations in wels catfish (*Silurus glanis*) from the Danube River by ICP-MS. *Environmental Science and Pollution Research*, **22**(5), 3820–3827.

Kelley, D. R., Schatz, M. C., & Salzberg, S. L. (2010). Quake: quality-aware detection and correction of sequencing errors. Genome Biology, 11(11), R116.

Kim, R., Emi, M., Tanabe, K., & Murakami, S. (2006). Role of the unfolded protein response in cell death. *Apoptosis*, **11**(1), 5–13.

Kragh-Hansen, U. (1981). Molecular aspects of ligand binding to serum albumin. *Pharmacological Reviews*, **33**(1), 17–53.

Krasnov, A., Afanasyev, S., & Oikari, A. (2007). Hepatic responses of gene expression in juvenile brown trout (*Salmo trutta lacustris*) exposed to three model contaminants applied singly and in combination. *Environmental Toxicology and Chemistry*, **26**(1), 100.

Krasnov, A., Koskinen, H., Pehkonen, P., Rexroad, C. E., Afanasyev, S., & Mölsä, H. (2005). Gene expression in the brain and kidney of rainbow trout in response to handling stress. *BMC Genomics*, **6**(1), 3.

Krumschnabel, G., Manzl, C., Berger, C., & Hofer, B. (2005). Oxidative stress, mitochondrial permeability transition, and cell death in Cu-exposed trout hepatocytes. *Toxicology and Applied Pharmacology*, **209**(1), 62–73.

Kumar, G., & Denslow, N. D. (2016). Gene Expression Profiling in Fish Toxicology: A Review (pp. 1–38). Springer International Publishing.

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, **10**(3), R25.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**(4), 357–359.

Langston, W., Chesman, B., Burt, G., Pope, N., & McEvoy, J. (2002). Metallothionein in liver of eels *Anguilla anguilla* from the Thames Estuary: an indicator of environmental quality? *Marine Environmental Research*, **53**(3), 263–293.

Li, B., Dewey, C. N.(2011). RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics*, **12**(1), 323.

Li, W., & Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**(13), 1658–1659.

Lin, M. T., & Beal, M. F. (2006). Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases. *Nature*, **443**(7113), 787–795.

Liu, S., Mansour, M. N., Dillman, K. S., Perez, J. R., Danley, D. E., Aeed, P. A., Simons, S.P., Lemotte, P.K., & Menniti, F.S. (2008). Structural basis for the catalytic mechanism of human phosphodiesterase 9. *Proceedings of the National Academy of Sciences of the United States of America*, **105**(36), 13309–14.

Liu, S., Wang, X., Sun, F., Zhang, J., Feng, J., Liu, H. Rajendran, K.V., Sun, L., Zhang, Y., Jiang, Y., Peatman, E., Kaltenboeck, L., Kucuktas, H., & Liu, Z. (2013). RNA-seq reveals expression signatures of genes involved in oxygen transport, protein synthesis, folding, and degradation in response to heat stress in catfish. *Physiological Genomics*, *45*(12).

Lu, X.J., Chen, J., Huang, Z.A., Zhuang, L., Peng, L.Z., & Shi, Y.H. (2012).

Influence of acute cadmium exposure on the liver proteome of a teleost fish, ayu (*Plecoglossus altivelis*). *Molecular Biology Reports*, **39**(3), 2851–2859.

Lu, M.W., Ngou, F.H., Chao, Y.M., Lai, Y.S., Chen, N.Y., Lee, F.Y., & Chiou, P. P. (2012). Transcriptome characterization and gene expression of *Epinephelus spp* in endoplasmic reticulum stress-related pathway during betanodavirus infection *in vitro*. *BMC Genomics*, **13**(1), 651.

Luza, S. C., & Speisky, H. C. (1996). Liver copper storage and transport during development: implications for cytotoxicity. *The American Journal of Clinical Nutrition*, **63**(5), 812S–20S.

Majumder, P. K., Mishra, N. C., Sun, X., Bharti, A., Kharbanda, S., Saxena, S., & Kufe, D. (2001). Targeting of protein kinase C delta to mitochondria in the oxidative stress response. *Cell Growth & Differentiation : The Molecular Biology Journal of the American Association for Cancer Research*, **12**(9), 465–70.

Masuoka, J., Hegenauer, J., Van Dyke, B. R., & Saltman, P. (1993). Intrinsic stoichiometric equilibrium constants for the binding of zinc(II) and copper(II) to the high affinity site of serum albumin. *The Journal of Biological Chemistry*, **268**(29), 21533–7.

McGeer, J. C., Szebedinszky, C., Gordon McDonald, D., & Wood, C. M. (2000). Effects of chronic sublethal exposure to waterborne Cu, Cd or Zn in rainbow trout 2: tissue specific metal accumulation. *Aquatic Toxicology*, **50**(3), 245–256.

Mehinto, A. C., Prucha, M. S., Colli-Dula, R. C., Kroll, K. J., Lavelle, C. M., Barber, D. S., Vuple, C.D., & Denslow, N. D. (2014). Gene networks and toxicity pathways induced by acute cadmium exposure in adult largemouth bass (*Micropterus salmoides*). *Aquatic Toxicology*, **152**, 186–194.

Miller, J. R. (1997). The role of fluvial geomorphic processes in the dispersal of heavy metals from mine sites. *Journal of Geochemical Exploration*, **58**(2-3), 101–118.

Minghetti, M., Schnell, S., Chadwick, M. A., Hogstrand, C., & Bury, N. R. (2014). A primary FIsh Gill Cell System (FIGCS) for environmental monitoring of river waters. *Aquatic Toxicology*, **154**, 184–192.

Misra, S., Zafarullah, M., Price-Haughey, J., & Gedamu, L. (1989). Analysis of stress-induced gene expression in fish cell lines exposed to heavy metals and heat shock. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, **1007**(3), 325–333.

Mori, T., Hiraka, I., Kurata, Y., Kawachi, H., Mano, N., Devlin, R. H., Nagoya, H., & Araki, K. (2007). Changes in hepatic gene expression related to innate immunity, growth and iron metabolism in GH-transgenic amago salmon (*Oncorhynchus masou*) by cDNA subtraction and microarray analysis, and serum lysozyme activity. *General and Comparative Endocrinology*, **151**(1), 42–54.

Naddy, R. B., Cohen, A. S., & Stubblefield, W. A. (2015). The interactive toxicity of cadmium, copper, and zinc to *Ceriodaphnia dubia* and rainbow trout (*Oncorhynchus mykiss*). *Environmental Toxicology and Chemistry*, **34**(4), 809–815.

Northcutt, A. J., Lett, K. M., Garcia, V. B., Diester, C. M., Lane, B. J., Marder, E., & Schulz, D. J. (2016). Deep sequencing of transcriptomes from the nervous systems of two decapod crustaceans to characterize genes important for neural circuit function and modulation. *BMC Genomics*, **17**(1), 868.

Okabe, H., Delgado, E., Lee, J.M., Yang, J., Kinoshita, H., Hayashi, H. Tsung, A., Behari, J., Beppu, T., Baba, H., & Monga, S.P. (2014). Role of leukocyte cell-derived chemotaxin 2 as a biomarker in hepatocellular carcinoma. *PLoS ONE*, **9**(6), e98817.

Ono, H., Ishii, K., Kozaki, T., Ogiwara, I., Kanekatsu, M., & Yamada, T. (2015). Removal of redundant contigs from de novo RNA-seq assemblies via homology search improves accurate detection of differentially expressed genes. *BMC Genomics*, **16**, 1031.

Osuna-Jiménez, I., Williams, T. D., Prieto-Álamo, M.-J., Abril, N., Chipman, J. K., & Pueyo, C. (2009). Immune- and stress-related transcriptomic responses of *Solea senegalensis* stimulated with lipopolysaccharide and copper sulphate using heterologous cDNA microarrays. *Fish & Shellfish Immunology*, **26**(5), 699–706.

Padmini, E., & Usha Rani, M. (2008). Impact of seasonal variation on HSP70 expression quantitated in stressed fish hepatocytes. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, **151**(3), 278–285.

Padmini, E., & Vijaya Geetha, B. (2009). Impact of season on liver mitochondrial oxidative stress and the expression of HSP70 in grey mullets from contaminated estuary. *Ecotoxicology*, **18**(3), 304–311.

Palermo, F. F., Risso, W. E., Simonato, J. D., & Martinez, C. B. R. (2015). Bioaccumulation of nickel and its biochemical and genotoxic effects on juveniles of the neotropical fish *Prochilodus lineatus*. *Ecotoxicology and Environmental Safety*, **116**, 19–28.

Paris, J. R., King, R. A., & Stevens, J. R. (2015). Human mining activity across the ages determines the genetic structure of modern brown trout (*Salmo trutta* L.) populations. *Evolutionary Applications*, **8**(6), 573–585.

Paquin, P. R., Santore, R. C., Wu, K. B., Kavvadas, C. D., & Di Toro, D. M. (2000). The biotic ligand model: a model of the acute toxicity of metals to aquatic life. *Environmental Science & Policy*, *3*, 175–182.

Parsell, D. A., & Lindquist, S. (1993). The function of heat-shock proteins in stress tolerance: degradation and reactivation of damaged proteins. *Annual Review of Genetics*, **27**(1), 437–496.

Pierron, F., Bourret, V., St-Cyr, J., Campbell, P. G. C., Bernatchez, L., &

Couture, P. (2009). Transcriptional responses to environmental metal exposure in wild yellow perch (*Perca flavescens*) collected in lakes with differing environmental metal concentrations (Cd, Cu, Ni). *Ecotoxicology*, **18**(5), 620–631.

Playle, R. C. (1998). Modelling metal interactions at fish gills. *Science of the Total Environment*, **219**(2), 147–163.

Rao, R. V, & Bredesen, D. E. (2004). Misfolded proteins, endoplasmic reticulum stress and neurodegeneration. *Current Opinion in Cell Biology*, **16**(6), 653–62.

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**(1), 139–140.

Roch, M., McCarter, J. A., Matheson, A. T., Clark, M. J. R., & Olafson, R. W. (1982). Hepatic metallothionein in rainbow trout (*Salmo gairdneri*) as an indicator of metal pollution in the Campbell River system. *Canadian Journal of Fisheries and Aquatic Sciences*, **39**(12), 1596–1601.

Roesijadi, G. (1994). Metallothionein induction as a measure of response to metal exposure in aquatic animals. *Environmental Health Perspectives*, **102** (Suppl 12), 91–5.

Ron, D., & Walter, P. (2007). Signal integration in the endoplasmic reticulum unfolded protein response. *Nature Reviews Molecular Cell Biology*, **8**(7), 519–529.

Rosabal, M., Pierron, F., Couture, P., Baudrimont, M., Hare, L., & Campbell, P. G. C. (2015). Subcellular partitioning of non-essential trace metals (Ag, As, Cd, Ni, Pb, and Tl) in livers of American (*Anguilla rostrata*) and European (*Anguilla anguilla*) yellow eels. *Aquatic Toxicology*, **160**, 128–141.

Salomons, W., & Eagle, A. (1990). Hydrology, sedimentology and the fate and distribution of copper in mine-related discharges in the fly river system, Papua New Guinea. *Science of The Total Environment*, **97-98**, 315–334.

Sanchez, W., Palluel, O., Meunier, L., Coquery, M., Porcher, J.-M., & Aït-Aïssa, S. (2005). Copper-induced oxidative stress in three-spined stickleback: relationship with hepatic metal levels. *Environ. Toxicol. Pharmacol.* **19**(1): 177:83.

Sanchez-Dardon, J., Voccia, I., Hontela, A., Chilmonczyk, S., Dunier, M., Boermans, H., Blakley, B. and Fournier, M. (1999), Immunomodulation by heavy metals tested individually or in mixtures in rainbow trout (*Oncorhynchus mykiss*) exposed in vivo. *Environmental Toxicology and Chemistry*, **18**: 1492–1497.

Santos, E. M., Ball, J. S., Williams, T. D., Wu, H., Ortega, F., van Aerle, R., Katsiadaki, I., Falciani, F., Viant, M.R., Chipman, J.K., & Tyler, C.R. (2010). Identifying health impacts of exposure to copper using transcriptomics and metabolomics in a fish model. *Environmental Science & Technology*, **44**(2),

820–826.

Schoville, S. D., Barreto, F. S., Moy, G. W., Wolff, A., & Burton, R. S. (2012). Investigating the molecular basis of local adaptation to thermal stress: population differences in gene expression across the transcriptome of the copepod *Tigriopus californicus*. *BMC Evolutionary Biology*, **12**(1), 170.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, **31**(19), 3210–3212.

Sokolova, I. M., Frederich, M., Bagwe, R., Lannig, G., & Sukhotin, A. A. (2012). Energy homeostasis as an integrative tool for assessing limits of environmental stress tolerance in aquatic invertebrates. *Marine Environmental Research*, **79**, 1–15.

Song, Y.-F., Huang, C., Shi, X., Pan, Y.-X., Liu, X., & Luo, Z. (2016). Endoplasmic reticulum stress and dysregulation of calcium homeostasis mediate Cu-induced alteration in hepatic lipid metabolism of javelin goby *Synechogobius hasta*. *Aquatic Toxicology*, **175**, 20–29.

Srikanth, K., Pereira, E., Duarte, A. C., & Ahmad, I. (2013). Glutathione and its dependent enzymes' modulatory responses to toxic metals and metalloids in fish—a review. *Environmental Science and Pollution Research*, **20**(4), 2133–2149.

Sundaram, A., Tengs, T. & Grimholt, U. 2017. Issues with RNA-seq analysis in non-model organisms: A salmonid example. *Dev. Comp. Immunol.* (**17**)30083-6.

Sweet, L.I. & Zelikof, J.T. 2001. Toxicology and immunotoxicology of mercury: a comparative review in fish and humans. *Journal of Toxicology and Environmental Health, Part B*, **4**(2), 161–205.

Tadiso, T. M., Krasnov, A., Skugor, S., Afanasyev, S., Hordvik, I., & Nilsen, F. (2011). Gene expression analyses of immune responses in Atlantic salmon during early stages of infection by salmon louse (*Lepeophtheirus salmonis*) revealed bi-phasic responses coinciding with the copepod-chalimus transition. *BMC Genomics*, **12**, 141.

Tallant, C., Marrero, A., & Gomis-Rüth, F. X. (2010). Matrix metalloproteinases: Fold and function of their catalytic domains. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, **1803**(1), 20–28.

Tamás, M. J., Sharma, S. K., Ibstedt, S., Jacobson, T., & Christen, P. (2014). Heavy metals and metalloids as a cause for protein misfolding and aggregation. *Biomolecules*, **4**(1), 252–67.

Uren Webster, T. M., Bury, N., van Aerle, R., & Santos, E. M. (2013). Global transcriptome profiling reveals molecular mechanisms of metal tolerance in a chronically exposed wild population of brown trout. *Environmental Science & Technology*, **47**(15), 8869–77.

Utgikar, V. P., Chaudhary, N., Koeniger, A., Tabak, H. H., Haines, J. R., & Govind, R. (2004). Toxicity of metals and metal mixtures: analysis of concentration and time dependence for zinc and copper. *Water Research*, **38**(17), 3651–3658.

van Heerden, D., Vosloo, A., & Nikinmaa, M. (2004). Effects of short-term copper exposure on gill structure, metallothionein and hypoxia-inducible factor-1α (HIF-1α) levels in rainbow trout (*Oncorhynchus mykiss*). *Aquatic Toxicology*, **69**(3), 271–280.

Valavanidis, A., Vlahogianni, T., Dassenakis, M., & Scoullos, M. (2006). Molecular biomarkers of oxidative stress in aquatic organisms in relation to toxic environmental pollutants. *Ecotoxicology and Environmental Safety*, **64**(2), 178–189.

Vembar, S. S., & Brodsky, J. L. (2008). One step at a time: endoplasmic reticulum-associated degradation. *Nature Reviews Molecular Cell Biology*, **9**(12), 944–957.

Wagner, A. (2005). Energy constraints on the evolution of gene Expression. *Molecular Biology and Evolution*, **22**(6), 1365–1374.

Waiho, K., Fazhan, H., Shahreza, M. S., Moh, J. H. Z., Noorbaiduri, S., Wong, L. L., Sinnasamy, S., & Ikhwanuddin, M. (2017). Transcriptome analysis and differential gene expression on the testis of orange mud crab, *Scylla olivacea*, during sexual maturation. *PLOS ONE*, **12**(1), e0171095.

Warnes, G., Bolker, B., Bonebakker, L., & Gentleman, R. (2009). gplots: Various R programming tools for plotting data. *R Package Version*.

Walter, W., Sánchez-Cabo, F., & Ricote, M. (2015). GOplot: an R package for visually combining expression data with functional analysis. *Bioinformatics*, **31**(17), 2912–2914.

Wang, Y., Ferrari, M. C. O., Hoover, Z., Yousafzai, A. M., Chivers, D. P., & Niyogi, S. (2014). The effects of chronic exposure to environmentally relevant levels of waterborne cadmium on reproductive capacity and behaviour in fathead minnows. *Archives of Environmental Contamination and Toxicology*, **67**(2), 181–191.

Whiting, P. J., Matisoff, G., Fornes, W., & Soster, F. M. (2005). Suspended sediment sources and transport distances in the Yellowstone River basin. *Geological Society of America Bulletin*, **117**(3), 515.

Wilkinson, B., & Gilbert, H. F. (2004). Protein disulfide isomerase. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, **1699**(1), 35–44.

Wood, C. M. 2012. An introduction to metals in fish physiology and toxicology: Basic principles. In: Wood, C., Farrell, A.P., and Brauner C.J. eds., *Homeostasis and Toxicology of Essential Metals*. Oxford: Elsevier Academic Press, pp. 2-40.

Xu, C., Evensen, Ø., & Munang'andu, H. (2015). De novo assembly and

transcriptome analysis of Atlantic salmon macrophage/dendritic-like TO cells following type I IFN treatment and Salmonid alphavirus subtype-3 infection. *BMC Genomics*, **16**(1), 96.

Yeşilbudak, B., & Erdem, C. (2014). Cadmium accumulation in gill, liver, kidney and muscle tissues of common carp, *Cyprinus carpio*, and Nile Tilapia, *Oreochromis niloticus*. *Bulletin of Environmental Contamination and Toxicology*, **92**(5), 546–550.

Zelikoff, J. T. (1993). Metal pollution-induced immunomodulation in fish. *Annual Review of Fish Diseases*, **3**, 305–325.

Zeng, F., Tee, C., Liu, M., Sherry, J. P., Dixon, B., Duncker, B. P., & Bols, N. C. (2014). The p53/HSP70 inhibitor, 2-phenylethynesulfonamide, causes oxidative stress, unfolded protein response and apoptosis in rainbow trout cells. *Aquatic Toxicology*, **146**, 45–51.

Zheng, G.-H., Liu, C.-M., Sun, J.-M., Feng, Z.-J., & Cheng, C. (2014). Nickel-induced oxidative stress and apoptosis in *Carassius auratus* liver by JNK pathway. *Aquatic Toxicology*, **147**, 105–111.

Zheng, J.-L., Yuan, S.-S., Wu, C.-W., & Li, W.-Y. (2016). Chronic waterborne zinc and cadmium exposures induced different responses towards oxidative stress in the liver of zebrafish. *Aquatic Toxicology*, **177**, 261–268

Supporting Information

Natural populations of chronically exposed metal-tolerant trout display minimal hepatic transcriptional responses to an acute metal exposure

This supporting information contains:

Table S1. Summary statistics of reads for each of the 24 samples.

Table S2. Differentially expressed (DE) genes upregulated in naïve-control fish.

Table S3. Differentially expressed (DE) genes downregulated in naïve-control fish.

Table S4. Significantly enriched Gene Ontology (GO) terms and KEGG pathways for differentially expressed genes in naïve-exposed fish.

Table S5. Differentially expressed (DE) genes over-expressed in tolerant-exposed fish in comparison to naïve-exposed fish.

Table S1. Summary statistics of reads for each of the 24 samples, including number of raw reads, number of reads after cleaning and normalisation and number of transcripts recovered. Total and Averages for overall raw reads, cleaned reads and transcripts.

Population	*Cohort*	*Sample ID*	*No. of raw reads*	*No. of clean & normalised reads*	*No. of transcripts*
Naïve	Control	Naïve-control 01	20,234,233	12,928,602	6,783,262
Naïve	Control	Naïve-control 02	22,889,674	15,144,105	8,394,201
Naïve	Control	Naïve-control 03	25,611,269	16,099,056	8,675,883
Naïve	Control	Naïve-control 04	31,982,721	20,184,199	10,725,806
Naïve	Control	Naïve-control 05	10,503,946	6,637,855	3,585,540
Naïve	Control	Naïve-control 06	13,881,568	8,927,724	4,606,515
Tolerant	Control	Tolerant-control 01	8,901,794	5,037,008	2,559,159
Tolerant	Control	Tolerant-control 02	37,450,985	23,813,812	13,003,185
Tolerant	Control	Tolerant-control 03	10,004,536	6,490,368	3,648,511
Tolerant	Control	Tolerant-control 04	7,775,802	4,975,807	2,815,167
Tolerant	Control	Tolerant-control 05	4,784,067	2,993,227	1,730,449
Tolerant	Control	Tolerant-control 06	7,515,650	4,598,774	2,508,232
Naïve	Exposed	Naïve-exposed 01	26,456,391	17,201,208	8,880,932
Naïve	Exposed	Naïve-exposed 01	33,456,741	21,001,184	11,226,341
Naïve	Exposed	Naïve-exposed 01	15,984,482	10,254,792	5,406,001
Naïve	Exposed	Naïve-exposed 01	21,115,259	13,331,935	6,561,265
Naïve	Exposed	Naïve-exposed 01	28,598,033	17,205,199	9,033,878
Naïve	Exposed	Naïve-exposed 01	4,601,393	2,846,383	1,419,179
Tolerant	Exposed	Tolerant-exposed 01	12,991,702	8,116,741	4,208,945
Tolerant	Exposed	Tolerant-exposed 02	31,033,901	18,809,650	10,172,994
Tolerant	Exposed	Tolerant-exposed 03	9,943,735	6,018,844	3,021,470
Tolerant	Exposed	Tolerant-exposed 04	32,889,493	20,797,269	11,380,200
Tolerant	Exposed	Tolerant-exposed 05	13,074,740	8,167,819	4,475,344
Tolerant	Exposed	Tolerant-exposed 06	10,918,932	6,550,817	3,569,671
		Total	442,601,047	278,132,378	148,392,130
		Average	18,441,710	11,588,849	6,183,005
		STD	10,293,468	6,511,138	3,507,301
		SEM	2,101,145	1,329,081	715,925

Table S2. Differentially expressed (DE) genes upregulated in naïve-control fish. Results shown for annotated genes significantly expressed at two-fold DE, present in 50% of samples at an FDR <0.05. Information includes transcript ID, uniprot annotation, official gene symbol and gene name, log-fold change (logFC), log counts-per-million (logCPM) and FDR corrected p-value.

Transcript ID	*Uniprot annotation*	*Official Gene Symbol*	*Gene name*	*logFC*	*logCPM*	*FDR*
TRINITY_DN70365_c0_g2	LECT2_BOVIN	*lect2*	leukocyte cell derived chemotaxin 2(LECT2)	9.96	8.39	0.00E+00
TRINITY_DN82044_c22_g1	TCB2_CAEBR	*N.A*	Transposable element Tcb2 transposase	8.76	8.16	0.00E+00
TRINITY_DN50027_c0_g1	TAR1_YEAST	*tar1*	Tar1p(TAR1)	6.89	1.89	5.00E-02
TRINITY_DN83595_c2_g5	UBP5_SCHPO	*ubp5*	ubiquitin C-terminal hydrolase Ubp5(ubp5)	6.71	2.05	4.00E-02
TRINITY_DN70910_c0_g3	WFDC2_RABIT	*wfdc2*	WAP four-disulfide core domain 2(WFDC2)	6.54	-0.23	1.00E-02
TRINITY_DN47970_c0_g1	ENPL_HUMAN	*hsp90b1*	heat shock protein 90 beta family member 1(HSP90B1)	5.56	0.21	0.00E+00
TRINITY_DN47295_c0_g1	LECT2_BOVIN	*lect2*	leukocyte cell derived chemotaxin 2(LECT2)	4.96	1.93	2.00E-02
TRINITY_DN39157_c0_g1	ISK1_ANGAN	*N.A*	Probable pancreatic secretory proteinase inhibitor	4.49	2.55	0.00E+00
TRINITY_DN34500_c0_g1	MFAP4_BOVIN	*mfap4*	microfibrillar associated protein 4(MFAP4)	4.16	8.96	0.00E+00
TRINITY_DN119081_c0_g1	C1QL2_HUMAN	*c1ql2*	complement C1q like 2(C1QL2)	3.92	4.45	2.00E-02
TRINITY_DN78376_c1_g4	GIMA4_MOUSE	*gimap4*	GTPase, IMAP family member 4(Gimap4)	3.9	1.78	1.00E-02
TRINITY_DN74452_c1_g1	MMP9_RABIT	*mmp9*	matrix metallopeptidase 9(MMP9)	3.84	1.93	3.00E-02
TRINITY_DN56907_c0_g1	PDE9A_RAT	*pde9a*	phosphodiesterase 9A(Pde9a)	3.8	1.58	3.00E-02
TRINITY_DN84605_c1_g1	TLR5_MOUSE	*tlr5*	toll-like receptor 5(Tlr5)	3.77	4.92	4.00E-02

TRINITY_DN81044_c6_g2	TNF13_HUMAN	*tnfsf13*	tumor necrosis factor superfamily member 13(TNFSF13)	3.66	1.21	0.00E+00
TRINITY_DN10324_c0_g1	MYLK_RABIT	*mylk*	myosin light chain kinase(MYLK)	3.65	10.76	2.00E-02
TRINITY_DN78102_c0_g4	CPNE4_MOUSE	*cpne4*	copine IV(Cpne4)	3.54	0.13	0.00E+00
TRINITY_DN89357_c0_g1	ITLN_ONCMY	*itln*	Intelectin	3.52	2.07	3.00E-02
TRINITY_DN51710_c0_g1	ENPL_PIG	*hsp90b1*	heat shock protein 90 beta family member 1(HSP90B1)	3.47	0.85	0.00E+00
TRINITY_DN61731_c0_g1	FKB11_MOUSE	*fkbp11*	FK506 binding protein 11(Fkbp11)	3.38	5.09	0.00E+00
TRINITY_DN82791_c0_g3	MMP9_CANLF	*mmp9*	matrix metallopeptidase 9(MMP9)	3.29	1.79	3.00E-02
TRINITY_DN83334_c1_g1	HYOU1_DANRE	*hyou1*	hypoxia up-regulated 1(hyou1)	3.25	4.98	0.00E+00
TRINITY_DN65643_c0_g1	DNJC3_HUMAN	*dnajc3*	DnaJ heat shock protein family (Hsp40) member C3(DNAJC3)	3.21	5.59	0.00E+00
TRINITY_DN84942_c4_g1	Tm1	*tm1*	Tropomyosin 1(Tm1)	3.17	5.26	2.00E-02
TRINITY_DN81860_c0_g1	HYOU1_DANRE	*hyou1*	hypoxia up-regulated 1(hyou1)	3.16	5.56	0.00E+00
TRINITY_DN79378_c0_g1	DJB11_MOUSE	*dnajb11*	DnaJ heat shock protein family (Hsp40) member B11(Dnajb11)	3.1	1.19	1.00E-02
TRINITY_DN77412_c1_g2	CNPY2_HUMAN	*cnpy2*	canopy FGF signaling regulator 2(CNPY2)	3.07	4.19	0.00E+00
TRINITY_DN60871_c0_g1	PLPL9_HUMAN	*pla2g6*	85/88 kDa calcium-independent phospholipase A2	2.98	2.58	0.01
TRINITY_DN79378_c0_g2	DJB11_PONAB	*dnajb11*	DnaJ heat shock protein family (Hsp40) member B11(DNAJB11)	2.96	4.59	1.00E-02
TRINITY_DN51259_c0_g1	SAA_CANLF	*saa1*	serum amyloid A1(SAA1)	2.92	4.78	1.00E-02
TRINITY_DN86382_c0_g1	EHD1_ARATH	*ehd1*	EPS15 homology domain 1(EHD1)	2.76	2.05	2.00E-02
TRINITY_DN67219_c0_g1	MANF_BOVIN	*manf*	mesencephalic astrocyte derived neurotrophic factor(MANF)	2.76	5.06	0.00E+00
TRINITY_DN180253_c0_g1	CALR_MACFU	*calr*	calreticulin(CALR)	2.75	4.36	0.00E+00

TRINITY_DN83209_c2_g1	HPT_CEREL	*hp*	Haptoglobin	2.73	13.07	0.00E+00
TRINITY_DN82208_c5_g1	CALR_HUMAN	*calr*	calreticulin(CALR)	2.53	6.38	0.00E+00
TRINITY_DN31754_c0_g1	PDIA3_PONAB	*pdia3*	protein disulfide isomerase family A member 3(PDIA3)	2.48	2.39	1.00E-02
TRINITY_DN82825_c0_g1	RGF1B_HUMAN	*rasgef1b*	RasGEF domain family member 1B(RASGEF1B)	2.39	1.69	0.00E+00
TRINITY_DN84956_c0_g1	STT3A_BOVIN	*stt3a*	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3A	2.37	6.29	0.00E+00
TRINITY_DN7842_c0_g1	TRP3_PSEAM	*trp3*	Trypsinogen-like protein 3	2.36	4.57	0.00E+00
TRINITY_DN79617_c1_g1	WDR7_HUMAN	*wdr7*	WD repeat domain 7(WDR7)	2.31	4.19	0.00E+00
TRINITY_DN60209_c1_g1	GRP78_CHICK	*hspa70*	heat shock 70kDa protein 5 (glucose-regulated protein, 78kDa)(HSPA5)	2.28	8.8	0.00E+00
TRINITY_DN86998_c0_g1	MOGS_MOUSE	*mogs*	mannosyl-oligosaccharide glucosidase(Mogs)	2.28	3.11	2.00E-02
TRINITY_DN78460_c0_g1	AT2A2_CANLF	*atp2a2*	ATPase sarcoplasmic/endoplasmic reticulum Ca2+ transporting 2(ATP2A2)	2.22	3.13	0.00E+00
TRINITY_DN68955_c1_g4	GMPPB_DANRE	*gmppb*	GDP-mannose pyrophosphorylase B(gmppb)	2.22	3.07	2.00E-02
TRINITY_DN83804_c1_g1	ASPM_HUMAN	*aspm*	abnormal spindle microtubule assembly(ASPM)	2.14	1.28	2.00E-02
TRINITY_DN73439_c0_g1	ENPL_BOVIN	*hsp90b1*	heat shock protein 90 beta family member 1(HSP90B1)	2.14	9.32	0.00E+00
TRINITY_DN85506_c7_g2	STT3A_BOVIN	*stt3a*	STT3A, catalytic subunit of the oligosaccharyltransferase complex(STT3A)	2.1	3.84	2.00E-02
TRINITY_DN102031_c0_g1	SDF2L_HUMAN	*sdf2l1*	stromal cell derived factor 2 like 1(SDF2L1)	2.09	1.39	4.00E-02
TRINITY_DN83968_c0_g1	JUNB_CYPCA	*junb*	Transcription factor jun-B	2.08	5.8	5.00E-02
TRINITY_DN82872_c0_g2	PDIA4_HUMAN	*pdia4*	protein disulfide isomerase family A member 4(PDIA4)	2.06	6.84	0.00E+00
TRINITY_DN62329_c0_g1	FKB14_PONAB	*fkbp14*	FK506 binding protein 14(FKBP14)	2.03	1.62	2.00E-02
TRINITY_DN80126_c0_g1	CALR_RABIT	*calr*	Calreticulin (CALR)	2.01	9.03	1.00E-02
TRINITY_DN112262_c0_g1	PDIA6_HUMAN	*pdia6*	protein disulfide isomerase family A member 6(PDIA6)	2.01	1.21	2.00E-02

Table S3. Differentially expressed (DE) genes downregulated in naïve-control fish. Results shown for annotated genes significantly expressed at two-fold DE, present in 50% of samples at an FDR <0.05. Information includes transcript ID, uniprot annotation, official gene symbol and gene name, log-fold change (logFC), log counts-per-million (logCPM) and FDR corrected p-value.

Transcript ID	*Uniprot annotation*	*Official Gene Symbol*	*Gene name*	*logFC*	*logCPM*	*FDR*
TRINITY_DN48000_c0_g1	POL_WDSV	*gap-pro-pol*	Pr gag-pro-pol(gap-pro-pol)	-10.61	3.78	0.05
TRINITY_DN85561_c0_g1	KCRM_RAT	*ckm*	creatine kinase, M-type(Ckm)	-9.61	6.49	0.02
TRINITY_DN80928_c0_g1	MLRS_RAT	*mylpf*	myosin light chain, phosphorylatable, fast skeletal muscle(Mylpf)	-8.92	5.31	0.05
TRINITY_DN86124_c1_g1	HA1F_CHICK	*bf2*	Major histocompatibility complex class I antigen BF2(BF2)	-7.44	0.80	0.00
TRINITY_DN78034_c1_g1	QNR71_COTJA	*tm1*	Tropomyosin 1(Tm1)	-6.06	0.95	0.00
TRINITY_DN85520_c0_g1	GVIN1_MOUSE	*gm4070*	predicted gene 4070(Gm4070)	-5.65	2.24	0.04
TRINITY_DN84071_c2_g1	PHLB2_MOUSE	*phldb2*	pleckstrin homology like domain, family B, member 2(Phldb2)	-5.11	2.13	0.00
TRINITY_DN84388_c1_g3	Tm1	*tm1*	tropomyosin 1(Tm1)	-4.94	0.06	0.03
TRINITY_DN56596_c0_g1	CR3LB_DANRE	*creb3l3a*	cAMP responsive element binding protein 3-like 3a(creb3l3a)	-4.60	-0.20	0.01
TRINITY_DN20155_c0_g1	ACO12_HUMAN	*acot12*	acyl-CoA thioesterase 12(ACOT12)	-4.52	1.73	0.04
TRINITY_DN55304_c0_g1	G3P_DANRE	*gapdh*	glyceraldehyde-3-phosphate dehydrogenase(gapdh)	-3.26	0.01	0.00
TRINITY_DN83970_c1_g3	IP6K2_RABIT	*ip6k2*	inositol hexakisphosphate kinase 2(IP6K2)	-3.18	-0.06	0.02

TRINITY_DN84046_c4_g1	YI31B_YEAST	*yil082w-a*	gag-pol fusion protein(YIL082W-A)	-3.10	4.07	0.00
TRINITY_DN78560_c0_g1	GBRR3_HUMAN	*gabrr3*	gamma-aminobutyric acid type A receptor rho3 subunit (gene/pseudogene)(GABRR3)	-3.07	1.60	0.05
TRINITY_DN64017_c0_g5	ALBU1_SALSA	*alb1*	serum albumin 1(alb1)	-2.49	2.93	0.03
TRINITY_DN87890_c2_g4	TRIO_DANRE	*trioa*	trio Rho guanine nucleotide exchange factor a(trioa)	-2.43	0.60	0.01
TRINITY_DN48349_c0_g1	ALBU1_SALSA	*alb1*	serum albumin 1(alb1)	-2.42	0.79	0.04
TRINITY_DN85707_c0_g1	DUS4_RAT	*dusp4*	dual specificity phosphatase 4(Dusp4)	-2.35	1.48	0.01
TRINITY_DN119999_c0_g1	ATTY_RAT	*tat*	tyrosine aminotransferase(Tat)	-2.31	0.16	0.02
TRINITY_DN69450_c0_g1	K1C20_BOVIN	*krt20*	keratin 20(KRT20)	-2.28	1.23	0.03
TRINITY_DN76867_c0_g1	POL4_DROME	*gpnmb*	glycoprotein nmb(GPNMB)	-2.24	3.51	0.04
TRINITY_DN57878_c0_g1	F173A_MOUSE	*fam173a*	family with sequence similarity 173, member A(Fam173a)	-2.23	2.85	0.00
TRINITY_DN85130_c0_g1	CHPF2_HUMAN	*chpf2*	chondroitin polymerizing factor 2(CHPF2)	-2.03	2.00	0.01

Table S4. Significantly enriched Gene Ontology (GO) terms and KEGG pathways for differentially expressed genes in naïve-exposed fish obtained by analysis in DAVID. Data includes 24 GO Biological Process (BP); 2 GO Molecular Function (MF), 12 GO Cellular Component (CC) and 4 KEGG pathways. Information includes Type (BP, MF, CC or KEGG), the GO/KEGG ID, Term description, the genes involved, the fold-enrichment and the Benjamini corrected FDR of each enrichment.

Type	*GO/KEGG ID*	*Term Description*	*Genes*	*Fold Enrichment*	*Benjamini FDR*
GP BP	GO:0090200	positive regulation of release of cytochrome c from mitochondria	*mmp9, pla2g6*	48.92	8.59E-03
GP BP	GO:0006984	ER-nucleus signaling pathway	*hsp90b1, atp2a2, hspa5, calr*	31.80	1.55E-02
GP BP	GO:0090199	regulation of release of cytochrome c from mitochondria	*mmp9, pla2g6*	29.81	1.66E-02
GP BP	GO:0001836	release of cytochrome c from mitochondria	*mmp9, pla2g6*	25.44	2.35E-02
GP BP	GO:0030433	ER-associated ubiquitin-dependent protein catabolic process	*hsp90b1, sdf2l1*	21.04	2.34E-03
GP BP	GO:0006457	protein folding	*hsp90b1, dnajb11, fkbp14, pdia6, hspa5, pdia4, calr, fkbp11*	19.93	1.47E-10
GP BP	GO:0036503	ERAD pathway	*hsp90b1, sdf2l1*	18.00	3.56E-03
GP BP	GO:0030968	endoplasmic reticulum unfolded protein response	*hsp90b1, pdia6, hspa5, calr, dnajc3*	16.79	1.45E-02
GP BP	GO:0035967	cellular response to topologically incorrect protein	*hsp90b1, sdf2l1, pdia6, hspa5, calr, dnajc3*	16.35	4.94E-03
GP BP	GO:0034620	cellular response to unfolded protein	*hsp90b1, pdia6, hspa5, calr, dnajc3*	16.22	1.50E-02
GP BP	GO:0035966	response to topologically incorrect protein	*hsp90b1, sdf2l1, pdia6, hspa5, calr, dnajc3, manf*	15.39	1.79E-03
GP BP	GO:0006986	response to unfolded protein	*hsp90b1, pdia6, hspa5, calr, dnajc3, manf*	15.14	6.33E-03
GP BP	GO:0034976	response to endoplasmic reticulum stress	*hsp90b1, sdf2l1, pdia6, pla2g6, hspa5, pdia4, calr, dnajc3*	14.79	2.01E-07
GP BP	GO:0044712	single-organism catabolic process	*hsp90b1, upb1, sdf2l1, mmp9, pla2g6, pde9a*	4.36	7.59E-03

GP BP	GO:0051603	proteolysis involved in cellular protein catabolic process	*ubp5, hsp90b1, sdf2l1, hspa5, dnajc3, usp50*	4.34	1.79E-02
GP BP	GO:0044257	cellular protein catabolic process	*ubp5, hsp90b1, sdf2l1, hspa5, dnajc3, usp50*	4.14	2.29E-02
GP BP	GO:0043066	negative regulation of apoptotic process	*hsp90b1, mmp9, hspa5, dnajc3*	4.02	5.21E-02
GP BP	GO:0051336	regulation of hydrolase activity	*rasgef1bb, hsp90b1, dnajb11, mmp9, rasgef1b, wfdc18, hspa5*	3.73	9.25E-03
GP BP	GO:0030163	protein catabolic process	*ubp5, hsp90b1, sdf2l1, hspa5, dnajc3, usp50*	3.54	5.45E-02
GP BP	GO:0042981	regulation of apoptotic process	*hsp90b1, pdia3, gimap8, sdf2l1, mmp9, pla2g6, hspa5, dnajc3, calr*	3.53	9.18E-03
GP BP	GO:0043067	regulation of programmed cell death	*hsp90b1, pdia3, gimap8, sdf2l1, mmp9, pla2g6, hspa5, dnajc3, calr*	3.49	9.64E-03
GP BP	GO:0010941	regulation of cell death	*hsp90b1, pdia3, gimap8, sdf2l1, mmp9, pla2g6, hspa5, dnajc3, calr*	3.24	1.47E-02
GP BP	GO:0006508	proteolysis	*ubp5, hsp90b1, sdf2l1, mmp9, wfdc18, hspa5, dnajc3, usp50*	2.85	2.85E-02
GP BP	GO:0012501	programmed cell death	*hsp90b1, pdia3, gimap8, sdf2l1, mmp9, pla2g6, hspa5, dnajc3, calr*	2.60	5.45E-02
GO MF	GO:0051082	unfolded protein binding	*hsp90b1, dnajb11, hspa5, calr*	23.29	9.32E-05
GO MF	GO:0046790	virion binding	*hsp90b1*	116.45	4.73E-04
GO CC	GO:0071682	endocytic vesicle lumen	*hsp90b1, calr*	327.14	3.66E-01
GO CC	GO:0034663	endoplasmic reticulum chaperone complex	*hsp90b1, dnajb11, sdf2l1, pdia6, hspa5*	163.57	2.13E-12
GO CC	GO:0005788	endoplasmic reticulum lumen	*hyou1, hsp90b1, pdia3, dnajb11, sdf2l1, fkbp14, pdia6, hspa5, pdia4, dnajc3, calr*	39.65	2.43E-18
GO CC	GO:0005790	smooth endoplasmic reticulum	*hsp90b1, calr, dnajc3*	33.46	2.35E-01
GO CC	GO:0042470	melanosome	*hsp90b1, pdia3, pdia6, pdia4*	19.53	1.11E-05
GO CC	GO:0030496	midbody	*hsp90b1, rasgef1b, hspa5, aspm*	13.26	9.63E-04
GO CC	GO:0031012	extracellular matrix	*hsp90b1, mmp9, calr, mfap4*	10.39	7.43E-04

GO CC	GO:0005925	focal adhesion	*hsp90b1, pdia3, hspa5, calr*	4.76	2.07E-01
GO CC	GO:0005783	endoplasmic reticulum	*hyou1, hsp90b1, pdia3, dnajb11, gimap8, cnpy2, pdia6, pde9a, mogs, dnajc3, calr, manf*	4.65	3.53E-04
GO CC	GO:0005789	endoplasmic reticulum membrane	*hsp90b1, dgat1, sdf2l1, pdia6, mogs, fkbp11*	3.69	1.67E-01
GO CC	GO:0005615	extracellular space	*saa1, mmp9, pla2g6, tnfsf13, hspa5, calr, manf*	3.29	3.08E-01
GO CC	GO:0070062	extracellular exosome	*cpne4, pdia3, mmp9, upb1, pdia6, tnfsf13, mogs, calr, hsp90b1, hspa5, dnajc3, mfap4, wfdc2, mylk*	2.75	2.30E-03
KEGG	hsa04141	Protein processing in endoplasmic reticulum	*hsp90b1, pdia6, pdia4, dnajc3, calr*	27.05	2.17E-03
KEGG	pon04141	Protein processing in endoplasmic reticulum	*dnajb11, pdia3*	38.57	8.76E-01
KEGG	hsa04918	Thyroid hormone synthesis	*hsp90b1, pdia4*	27.72	8.56E-01
KEGG	bta04141	Protein processing in endoplasmic reticulum	*hsp90b1, stt3a*	26.09	7.87E-01

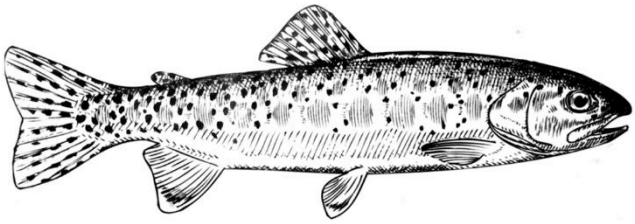
Table S5. Differentially expressed (DE) genes over-expressed in tolerant-exposed fish in comparison to naïve-exposed fish. Results shown for annotated genes significantly expressed at two-fold DE, present in 50% of samples at an FDR <0.05. Information includes transcript ID, uniprot annotation, official gene symbol and gene name, log-fold change (logFC), log counts-per-million (logCPM) and FDR corrected p-value.

Transcript ID	*Uniprot annotation*	*Official Gene Symbol*	*Gene name*	**logFC**	**logCPM**	**FDR**
TRINITY_DN84873_c2_g1	IGH	*igh*	immunoglobulin heavy locus(IGH)	10.72	5.48	7.96E-10
TRINITY_DN82034_c28_g8	CO4A1_DROME	*cg25c*	Collagen type IV(Cg25C)	8.36	1.83	0.00E+00
TRINITY_DN19382_c0_g1	YTX2_XENLA	*N.A*	Transposon TX1 uncharacterized 149 kDa protein	6.62	0.22	4.00E-02
TRINITY_DN61603_c1_g1	LORF2_MOUSE	$ORF2p$	LINE-1 retrotransposable element ORF2 protein	6.58	0.22	0.00E+00
TRINITY_DN87659_c3_g1	YTX2_XENLA	*N.A*	Transposon TX1 uncharacterized 149 kDa protein	6.19	2.19	0.00E+00
TRINITY_DN77976_c2_g1	COOA1_HUMAN	*col24a1*	collagen type XXIV alpha 1 chain(COL24A1)	6.18	1.06	0.00E+00
TRINITY_DN86557_c1_g2	HNRPC_XENLA	*hnrnpc.l*	heterogeneous nuclear ribonucleoprotein C (C1/C2) L homeolog(hnrnpc.L)	5.85	7.07	0.00E+00
TRINITY_DN85481_c0_g13	HMR1_PONPY	*mr1*	Major histocompatibility complex class I-related gene protein	5.76	3.8	5.00E-02
TRINITY_DN80523_c0_g1	INHBB_CHICK	*inhbb*	inhibin, beta B(INHBB)	5.74	2.78	0.00E+00
TRINITY_DN85307_c18_g17	STXB_SYNHO	*N.A*	Stonustoxin subunit beta	5.69	-0.36	2.00E-02
TRINITY_DN78557_c3_g1	TCB1_CAEBR	*N.A*	Transposable element Tcb2 transposase	5.64	-0.45	1.00E-02
TRINITY_DN67816_c1_g1	CASPE_HUMAN	*casp14*	caspase 14(CASP14)	5.55	-0.46	3.00E-02
TRINITY_DN118389_c0_g1	SPEG_DANRE	*spega*	SPEG complex locus a(spega)	5.53	-0.51	2.00E-02
TRINITY_DN27491_c0_g1	PUR2_CHICK	*gart*	phosphoribosylglycinamide formyltransferase, phosphoribosylglycinamide synthetase, phosphoribosylaminoimidazole synthetase(GART)	5.47	-0.55	1.00E-02
TRINITY_DN71247_c0_g1	RHES_RAT	*rasd2*	RASD family, member 2(Rasd2)	5.4	1.41	0.00E+00
TRINITY_DN59791_c0_g1	SPNS3_DANRE	*spns3*	spinster homolog 3 (Drosophila)(spns3)	5.35	1.45	4.00E-02
TRINITY_DN51357_c0_g1	CB080_HUMAN	*c2orf80*	chromosome 2 open reading frame 80(C2orf80)	5.21	-0.7	3.00E-02
TRINITY_DN85561_c0_g1	KCRM_RAT	*ckm*	creatine kinase, M-type(Ckm)	5.2	2.13	1.00E-02
TRINITY_DN176615_c0_g1	SPEG_DANRE	*spega*	SPEG complex locus a(spega)	5.02	-0.81	5.00E-02
TRINITY_DN80818_c3_g12	CHIA_HUMAN	*chia*	chitinase, acidic(CHIA)	4.99	4.89	4.00E-02
TRINITY_DN36055_c0_g1	POL3_DROME	*pol*	Retrovirus-related Pol polyprotein from transposon 17.6	4.51	2.09	0.00E+00

TRINITY_DN61139_c0_g1	ZN569_HUMAN	*znf569*	zinc finger protein 569(ZNF569)	4.37	1.45	0.00E+00
TRINITY_DN56619_c0_g2	PRRT4_RAT	*prrt4*	proline-rich transmembrane protein 4(Prrt4)	4.27	0.03	1.00E-02
TRINITY_DN34440_c0_g1	NEGR1_HUMAN	*negr1*	neuronal growth regulator 1(NEGR1)	4.07	-0.11	0.00E+00
TRINITY_DN84187_c2_g2	LAMA1_HUMAN	*lama1*	laminin subunit alpha 1(LAMA1)	3.99	3.84	0.00E+00
TRINITY_DN51758_c0_g1	ALBU2_SALSA	*loc100136922*	serum albumin 2(LOC100136922)	3.98	3.17	0.00E+00
TRINITY_DN82179_c0_g1	LAMA2_HUMAN	*lama2*	laminin subunit alpha 2(LAMA2)	3.91	0.48	1.00E-02
TRINITY_DN24593_c0_g1	HEBP2_HUMAN	*hebp2*	heme binding protein 2(HEBP2)	3.87	3.68	0.00E+00
TRINITY_DN72585_c0_g1	FKB15_HUMAN	*fkbp15*	FK506 binding protein 15(FKBP15)	3.83	-0.26	0.00E+00
TRINITY_DN82181_c3_g4	LORF2_MOUSE	*pol*	LINE-1 retrotransposable element ORF2 protein	3.82	1.25	0.00E+00
TRINITY_DN84071_c2_g1	PHLB2_MOUSE	*phldb2*	pleckstrin homology like domain, family B, member 2(Phldb2)	3.79	0.83	0.00E+00
TRINITY_DN83970_c1_g3	IP6K2_RABIT	*ip6k2*	inositol hexakisphosphate kinase 2(IP6K2)	3.74	0.24	1.00E-02
TRINITY_DN35554_c0_g1	ALBU2_SALSA	*loc100136922*	serum albumin 2(LOC100136922)	3.69	0.48	0.00E+00
TRINITY_DN76694_c0_g3	PTPRN_HUMAN	*ptprn*	protein tyrosine phosphatase, receptor type N(PTPRN)	3.66	2.2	0.00E+00
TRINITY_DN83810_c6_g2	VCAM1_CANLF	*vcam1*	vascular cell adhesion molecule 1(VCAM1)	3.61	0.79	1.00E-02
TRINITY_DN83970_c2_g1	IP6K2_RABIT	*ip6k2*	inositol hexakisphosphate kinase 2(IP6K2)	3.33	2.25	0.00E+00
TRINITY_DN64017_c0_g5	ALBU1_SALSA	*alb1*	serum albumin 1(alb1)	3.26	3.5	0.00E+00
TRINITY_DN72875_c0_g3	TNNI1_RABIT	*tnni1*	troponin I1, slow skeletal type(TNNI1)	3.24	0.84	3.00E-02
TRINITY_DN66967_c0_g1	RASF1_HUMAN	*rassf1*	Ras association domain family member 1(RASSF1)	3.23	0.8	3.00E-02
TRINITY_DN81987_c0_g1	ANGL8_MOUSE	*angptl8*	angiopoietin-like 8(Angptl8)	3.22	3.12	3.00E-02
TRINITY_DN61598_c0_g1	GLSK_HUMAN	*gls*	glutaminase(GLS)	3.22	-0.05	3.00E-02
TRINITY_DN78557_c5_g12	YTX2_XENLA	*N.A*	Transposon TX1 uncharacterized 149 kDa protein	3.2	0.22	1.00E-02
TRINITY_DN87478_c4_g1	TF29_SCHPO	*tf2-4*	retrotransposable element(Tf2-4)	3.16	0.04	0.00E+00
TRINITY_DN82598_c0_g1	NAR5_MOUSE	*art5*	ADP-ribosyltransferase 5(Art5)	3.15	0.87	0.00E+00
TRINITY_DN146819_c0_g1	PD5BA_XENLA	*pds5b.l*	PDS5 cohesin associated factor B L homeolog(pds5b.L)	3.1	-0.17	0.00E+00
TRINITY_DN47910_c0_g1	ALBU1_SALSA	*alb1*	serum albumin 1(alb1)	3.07	0.66	0.00E+00
TRINITY_DN60906_c0_g1	PA2GX_HUMAN	*pla2g10*	phospholipase A2 group X(PLA2G10)	3.06	3.69	1.00E-02
TRINITY_DN75888_c0_g1	AHI1_HUMAN	*ahi1*	Abelson helper integration site 1(AHI1)	3.05	0.56	0.00E+00
TRINITY_DN68403_c0_g1	ALBU2_SALSA	*loc100136922*	serum albumin 2(LOC100136922)	3.02	1.07	1.00E-02
TRINITY_DN84450_c1_g1	ITA9_HUMAN	*itga9*	integrin subunit alpha 9(ITGA9)	2.96	2.83	1.00E-02

TRINITY_DN55304_c0_g1	G3P_DANRE	*gapdh*	glyceraldehyde-3-phosphate dehydrogenase(gapdh)	2.91	-0.29	3.00E-02
TRINITY_DN88133_c0_g2	RDH12_BOVIN	*rdh12*	retinol dehydrogenase 12 (all-trans/9-cis/11-cis)(RDH12)	2.9	0.79	0.00E+00
TRINITY_DN48349_c0_g1	ALBU1_SALSA	*alb1*	serum albumin 1(alb1)	2.89	1.05	2.00E-02
TRINITY_DN83775_c3_g1	ESR2_SPAAU	*esr2*	type II estrogen receptor(LOC100534515)	2.87	3.12	0.00E+00
TRINITY_DN42215_c0_g1	PUR2_HUMAN	*gart*	phosphoribosylglycinamide formyltransferase, phosphoribosylglycinamide synthetase, phosphoribosylaminoimidazole synthetase(GART)	2.87	0.17	4.00E-02
TRINITY_DN87360_c0_g3	BCL6_HUMAN	*bcl6*	B-cell CLL/lymphoma 6(BCL6)	2.83	3.35	0.00E+00
TRINITY_DN33388_c0_g1	BHMT1_DANRE	*bhmt*	betaine-homocysteine methyltransferase(bhmt)	2.78	1.05	1.00E-02
TRINITY_DN28750_c0_g2	TBX3_HUMAN	*tbx3*	T-box 3(TBX3)	2.78	0.4	1.00E-02
TRINITY_DN87359_c1_g1	GBGT1_CHICK	*gbgt1*	globoside alpha-1,3-N-acetylgalactosaminyltransferase 1(GBGT1)	2.73	2.47	4.00E-02
TRINITY_DN62495_c0_g3	LSHR_CHICK	*lhcgr*	luteinizing hormone/choriogonadotropin receptor(LHCGR)	2.73	-0.22	3.00E-02
TRINITY_DN67162_c0_g1	GXLT1_MOUSE	*gxylt1*	glucoside xylosyltransferase 1(Gxylt1)	2.72	1.06	0.00E+00
TRINITY_DN58596_c0_g1	THMS1_HUMAN	*themis*	thymocyte selection associated(THEMIS)	2.72	4.41	4.00E-02
TRINITY_DN68346_c0_g1	DHR13_BOVIN	*dhrs13*	dehydrogenase/reductase 13(DHRS13)	2.67	2.68	1.00E-02
TRINITY_DN46942_c0_g1	SXT1_TAKPA	*psbp1*	Saxitoxin and tetrodotoxin-binding protein 1	2.67	10.84	0.00E+00
TRINITY_DN155637_c19_g1	APA11_ONCMY	*apoa-i-1*	apolipoprotein A-I-1(apoa-i-1)	2.64	5.54	1.00E-02
TRINITY_DN78421_c0_g1	TM163_BOVIN	*tmem163*	transmembrane protein 163(TMEM163)	2.63	0.68	5.00E-02
TRINITY_DN45195_c1_g1	PABP4_HUMAN	*pabpc4*	poly(A) binding protein cytoplasmic 4(PABPC4)	2.58	0.59	2.00E-02
TRINITY_DN76635_c0_g2	ITIH6_HUMAN	*itih6*	inter-alpha-trypsin inhibitor heavy chain family member 6(ITIH6)	2.54	2.72	2.00E-02
TRINITY_DN76730_c1_g1	PBLD_BOVIN	*pbld*	phenazine biosynthesis like protein domain containing(PBLD)	2.49	4.96	5.00E-02
TRINITY_DN87121_c1_g8	LEG1H_ONCMY	*leg1*	protein LEG1 homolog(leg1)	2.47	11.62	3.00E-02
TRINITY_DN81832_c3_g1	TF29_SCHPO	*tf2-4*	retrotransposable element(Tf2-4)	2.43	1.64	1.00E-02
TRINITY_DN78195_c0_g1	Tm1	*tm1*	Tropomyosin 1(Tm1)	2.37	3.99	0.00E+00
TRINITY_DN25204_c0_g1	THAP4_MOUSE	*thap4*	THAP domain containing 4(Thap4)	2.36	1.95	0.00E+00
TRINITY_DN53405_c1_g3	CP3AR_ONCMY	*cyp3a27*	Cytochrome P450 3A27	2.34	-0.01	4.00E-02
TRINITY_DN72984_c0_g1	ESR2_ORENI	*esr2*	type II estrogen receptor(LOC100534515)	2.34	1.71	1.00E-02
TRINITY_DN77537_c0_g1	PDS5B_CHICK	*pds5b*	PDS5, regulator of cohesion maintenance, homolog B (S. cerevisiae)(PDS5B)	2.32	1.22	0.00E+00

TRINITY_DN57878_c0_g1	F173A_MOUSE	*fam173a*	family with sequence similarity 173, member A(Fam173a)	2.29	2.79	0.00E+00
TRINITY_DN83732_c9_g3	ACBG2_XENLA	*acsbg2.l*	acyl-CoA synthetase bubblegum family member 2 L homeolog(acsbg2.L)	2.28	3.65	2.00E-02
TRINITY_DN24018_c0_g1	ALBU2_SALSA	*loc100136922*	serum albumin 2(LOC100136922)	2.28	2.14	0.00E+00
TRINITY_DN80501_c0_g1	LAYN_MOUSE	*layn*	layilin(Layn)	2.28	2.73	3.00E-02
TRINITY_DN72068_c0_g1	MYPOP_XENLA	*mypop.l*	Myb-related transcription factor, partner of profilin L homeolog(mypop.L)	2.26	1.92	0.00E+00
TRINITY_DN70606_c1_g1	REP15_HUMAN	*rep15*	RAB15 effector protein(REP15)	2.25	0.51	1.00E-02
TRINITY_DN82164_c3_g2	LAP2_RAT	*tmpo*	thymopoietin(Tmpo)	2.24	4.67	3.00E-02
TRINITY_DN80680_c0_g1	CNTN4_HUMAN	*cntn4*	contactin 4(CNTN4)	2.21	1.9	0.00E+00
TRINITY_DN41663_c0_g1	RET4B_ONCMY	*rbp4b*	Retinol-binding protein 4-B	2.2	0.71	5.00E-02
TRINITY_DN51825_c0_g1	RFTN2_HUMAN	*rftn2*	raftlin family member 2(RFTN2)	2.2	1.28	4.00E-02
TRINITY_DN69222_c0_g1	UD2A1_RAT	*ugt2a1*	UDP glucuronosyltransferase 2 family, polypeptide A1(Ugt2a1)	2.2	2.55	3.00E-02
TRINITY_DN43549_c0_g2	AKAP5_RAT	*akap5*	A-kinase anchoring protein 5(Akap5)	2.18	0.77	5.00E-02
TRINITY_DN85253_c1_g10	BT1A1_HUMAN	*btn1a1*	butyrophilin subfamily 1 member A1(BTN1A1)	2.17	3.13	0.00E+00
TRINITY_DN62049_c0_g1	GLYM_BOVIN	*shmt2*	serine hydroxymethyltransferase 2(SHMT2)	2.17	1.85	5.00E-02
TRINITY_DN82143_c0_g1	HCE1_ORYLA	*hce*	protease(hce)	2.16	1.01	3.00E-02
TRINITY_DN80574_c1_g3	ANXA2_MOUSE	*anxa2*	annexin A2(Anxa2)	2.09	3.66	0.00E+00
TRINITY_DN64826_c1_g1	LORF2_HUMAN	*loc100128274*	LINE-1 retrotransposable element ORF2 protein	2.05	0.98	1.00E-02
TRINITY_DN82429_c6_g3	TAP2_RAT	*tap2*	transporter 2, ATP binding cassette subfamily B member(Tap2)	2.04	6.53	0.00E+00
TRINITY_DN63171_c4_g1	LIPO_RHIMB	*n.a*	Lipocalin	2.03	10.06	0.00E+00



Chapter VII

GENERAL DISCUSSION

'The trout is no single, common, identical, definite, determined and measurable fish, but rather ten thousand tantalizing, distinct and different devils'

C.J. Holmes, The Tarn and the Lake: Thoughts on Life in the Italian Renaissance (London, 1913)

The objective of this thesis was to assess the relative contribution of genetic local adaptation and phenotypic plasticity in conferring metal-tolerance observed in several brown trout populations in southwest England. To this end, a series of experiments to investigate both the patterns and processes contributing to metal tolerance were undertaken. Below, a synthesis of these results, their strengths and their shortcomings is discussed.

7.1 EVIDENCING LOCAL ADAPTATION

Although investigating local adaptation is crucial in understanding the processes of evolution, providing evidence for it has, to date, proven extremely challenging. It has been advocated that an assessment of local adaptation should employ multifaceted approaches (Chevin *et al.* 2010) and it has been stressed that experiments should explicitly address hypotheses about the role of particular ecological and genetic factors that might promote or hinder local adaptation (Kawecki & Ebert 2004). Utilising genetic and genomic techniques, accompanied by experimental evidence, inarguably provides the strongest evidence for local adaptation. Together, such investigations inform better evolutionary predictions of local adaptation and an improved understanding of how organisms maintain themselves in the face of environmental change.

Using putatively neutral microsatellites (Chapter 2), strong demographic patterns associated with environmental disturbances as a result of mining activity were illustrated, with metal-impacted trout demonstrating low genetic diversity, population bottlenecks and marked genetic differentiation compared to trout from clean rivers. Moreover, the timing of these demographic changes were correlated with increased periods of mining intensity in the region. Together, this evidence pointed to the wide-ranging demographic effects that mining had (and continues to have) on these trout populations, and also strongly hinted at rapid local adaptation in metal-tolerant populations. However, interpretation of the results was limited to being reflective of neutral processes driven by intensive mining and contemporary metal pollution. Indeed, at its first submission, reviewers highlighted that we had not demonstrated local adaptation: ‘I agree that the authors have found patterns

of genetic structure that could be associated with mining activity, however, they have no evidence to suggest that local adaptation has occurred.' It was recommended that if different physiological mechanisms of metal tolerance were to be interrogated then conclusions of adaptive differentiation between populations could be evidenced. Furthermore, it was suggested that genomic data could be used to identify patterns of local adaptation by conducting a genome scan for loci under selection among the different populations. This leads on to the other investigations that the thesis has developed; namely, using multi-faceted approaches in order to provide more compelling support of local adaptation (where it appears to have occurred). Concurrently, through consideration of the results of additional research in this thesis, a robust assessment of the role of phenotypic plasticity in brown trout could be made.

Combining evidence through multiple investigative routes can significantly enlighten the actions of local adaptation. Populations of a bottom-dwelling fish, the tomcod (*Microgadus tomcod*) are known to have adapted to polychlorinated biphenyls (PCBs) in the Hudson River, New York. Evidence for this originated by marrying up the results of several varied investigations using a combination of gene expression data, exposure experiments and genetics. Early studies indicated that ecotoxicological biomarkers such as cytochrome p450 (CYP1A1) showed significantly higher gene expression in populations inhabiting the Hudson (Wirgin *et al.* 1994; Roy *et al.* 2002), yet when adult tomcod from the Hudson were chemically treated with PCBs after a week of depuration, no such induction was observed (Wirgin *et al.* 1992; Courtenay *et al.* 1999). Further analysis investigated physiological mechanisms of PCB handling, such as regulation of PCBs in hepatic tissue (Fernandez *et al.* 2004) and it was apparent that Hudson-river tomcod show a 100-fold reduction in sensitivity to early life-stage toxicity to PCBs (Wirgin & Chambers 2006). Ultimately, the underlying mechanism of this tolerance was evidenced through genetic techniques (Wirgin *et al.* 2011); sequencing of the aryl hydrocarbon receptor (AHR) gene, which was implicated in potential adaptation, identifying a 6bp deletion conferring PCB resistance in these fish (Wirgin *et al.* 2011). Similarly, local adaptation to pollutants (PCBs and PAHs) has been demonstrated in multiple populations of the Atlantic killifish

(*Fundulus heteroclitus*); this evidence, arising from a decade of complementary research was reviewed by Whitehead *et al.* (2017). Recently, both whole-genome sequencing data and RNA-seq data were used to demonstrate local adaptation across multiple control and exposed populations of the species (Reid *et al.* 2016).

The overall aim of the physiology experiment presented here was to explore the underlying mechanisms (if any) of metal-tolerance (Chapter 5 and Chapter 6). The purpose of these investigations was two-fold. Firstly, to assess whether metal-impacted trout really are adapted, and if so, the potential physiological mechanisms underlying such adaptation; and secondly, if such mechanisms exist, to what extent do they represent genetic local adaptation, rather than processes indicative of phenotypic plasticity. Evidence of acclimation (or phenotypic plasticity) to metal exposure in fish is common; for example, the seminal study by Laurén & McDonald (1987) showed significant physiological alterations in trout related to copper exposure over time. The fact that these plastic responses exist is due to the presence of inherent metal-handling mechanisms, such as metal-specific transporters and metal-storage proteins. Therefore, in comparison to 'biologically-unfamiliar' man-made chemical pollutants, such as polychlorinated biphenyls (PCBs), where such physiological acclimation is less likely, investigating the potential for phenotypic plasticity in conferring tolerance to metals is even more necessary.

The processes observed in Chapter 5, such as the accumulation of metals over the lifetimes of metal-tolerant trout and the inherent differences in osmoregulation, are indicative of local adaptation. This evidence was perhaps most well demonstrated by the obvious intrinsic differences in ammonia excretion between metal-tolerant and naïve fish. The likelihood that phenotypic plasticity is acting alone in driving such mechanisms would require these physiological changes to have arisen within the lifetime of an individual metal-tolerant trout. Resident trout of metal-impacted rivers emerge from eggs that have developed in metal-polluted environments and salmonid eggs are particularly vulnerable to metal toxicity (Finn 2007). The likelihood that metal-handling physiological changes occur *during* this time, and then are sustained through to maturity, would require significant energy and resources in the form

of constitutive upregulation of a suite of genes. This is also unlikely given the costs and limits of plasticity. Although we focused only on liver tissue, hepatic gene expression differences between exposed and control fish from metal-impacted trout populations showed no such patterns (Chapter 6). Even more striking was the fact that the upregulation of genes implicated in the in-river persistence of trout from the River Hayle (Uren-Webster *et al.* 2013) were not frontloaded (i.e. constitutively expressed), nor did they reanimate on re-exposure to metals. This is not to say that these genes are not important in allowing these fish to tolerate metals in-river, but they likely represent a 'second-line' of defence.

Finally, through investigating these myriad physiological processes, we put forward a hypothesis at the end of Chapter 5, which it is important to highlight here in the context of the other research chapters. Specifically, we propose that, together, the physiological patterns observed in metal-tolerant fish suggest that regulatory mechanisms exist for compensating for essential ion loss at the gill site. The gene ontology of outlier loci identified by the RADseq experiment (Chapter 4) showed that a number of genes potentially regulating osmoregulation and cytoskeleton gill structure are under selection. Regulation of osmoregulation therefore is likely a fundamental adaptive trait that exists in these populations. By increasing the uptake of essential ions, a higher amount of metals are then permitted entry into the bodies of the trout. In metal-naïve fish, these higher concentrations could be fatal, but metal-adapted trout do not appear to show any signs of reduced health. Although the liver represents a key organ involved in detoxification and excretion of metals, no genes were significantly upregulated on re-exposure (to admittedly) lower concentrations of metals (Chapter 6). This again suggests genetic traits underlying metal regulation do exist, further supporting the existence of local adaptation. Other loci identified as under selection (Chapter 4) may regulate such internal metal-handling processes, especially those involved in regulation of polyamine levels, DNA damage repair, cellular metabolism and translation regulation.

7.2 TRANSITION FROM GENETIC MARKERS TO GENOMICS

Neutral genetic markers have been the marker of choice for decades, but their original selection for neutrality seriously limits the potential for researchers to narrow down findings based on these loci to specific regions of a genome contributing to local adaptation. The problem of confidently evidencing local adaptation was apparent in Chapter 2 (microsatellite analysis) as the research described *patterns,* but by its analytical nature could not describe any *processes* contributing to local adaptation. Despite numerous and varied tests for identifying selection in the microsatellite loci used to characterise trout (Chapter 2), none showed any evidence that selection was operating in metal-impacted populations. There are of course several reasons for this. A panel of 23 microsatellite loci, whilst having higher statistical power than a previous study by Durrant *et al.* (2011) (which used 7 microsatellite loci), is still far fewer loci than may be needed to detect selection. On average, brown trout have approximately 40 chromosomes (Gharbi *et al.* 2006; Leitwein *et al.* 2017), thus, the number of loci used do not provide enough coverage for even one locus per chromosome. Further, we do not know where these loci are positioned within the genome. In a relatively crude assessment of this, the microsatellite loci were blasted against the *Salmo salar* genome; the results showed that the 23 loci aligned to just 16 of the 29 chromosomes, and five loci aligned multiple times to the same chromosome.

Furthermore, despite attempts to use microsatellite markers potentially linked to selection (MHC- and EST-linked loci), no evidence of selection was detected in the microsatellite loci used in this thesis. Marker choice played an important role here; indeed, we could have used other markers, such as those known to be linked to immune-relevant ESTs (Tonteri *et al.* 2008), but this demands *a priori* knowledge of the regions of the genome expected to be under selection – it's a long-shot. In the current study, the role of metallothioneins in promoting metal-tolerance was expected, given their protective role against metal toxicity (Hamilton & Mehrle 1986), and also

based on the results of a previous study of metal-tolerance in Hayle fish (Uren-Webster *et al.* 2013). The development of microsatellite markers in potential linkage with the two forms of metallothionein in salmonids (metA and metB) was investigated (results not included in this thesis), and all 700 samples used in Chapter 2 were screened for these markers. Unfortunately, even with this *a priori* knowledge of potential genes under selection, these loci were not shown to be under selection.

Reduced-representation sequencing, such as that offered by RADseq has proven to be a popular method in overcoming the limitations of neutral genetic markers. Indeed, in 2010, the RADseq method was highlighted in the *Science* feature 'Breakthrough of the Year' (DOI: 10.1126/science.330.6011.1608-b). Such genome-wide techniques have allowed researchers to begin to explore the minutiae of local adaptation, or 'adaptation genomics' (Stapely *et al.* 2010; Radwan & Babik 2012), and the higher resolution data generated by genomics has provided researchers with the ability to interrogate evolutionary forces at a much more detailed level. For example, in a study of the phylogeography of the crucian carp (*Carassius carassius*), RADseq elucidated finer population structure and stronger patterns of IBD than microsatellites, despite using only a proportion of the total samples analysed with microsatellites (Jeffries *et al.* 2016). As demonstrated in this thesis, the use of RADseq significantly revealed the role of local adaptation in metal-tolerant brown trout populations (Chapter 4), which could previously only be hinted at by exploring patterns in neutral microsatellite data (Chapter 2).

At the time of writing, a heated debate has emerged on the ability of RADseq to accurately detect local adaptation. Some researchers have argued that because estimates of linkage disequilibrium (LD) have not been ascertained for many species in studies that use RADseq, regions of the genome associated with local adaptation may be misinterpreted, or missed entirely (Lowry *et al.* 2017a). Responses to 'Breaking RAD' from two other groups have argued that a) RADseq can, and has been shown to, identify regions associated with local adaptation, which in some cases have been strengthened by independent data (McKinney *et al.* 2017) and b) that whilst LD should be considered, RADseq offers much greater resolution than could

previously be detected (as shown in this thesis by comparing the findings of Chapter 2 and Chapter 4) and, currently, alternative genome-wide methods of assessing local adaptation also have their limitations (e.g. Chapter 6 of this thesis) (Catchen *et al.* 2017). Lowry and colleagues have responded that the best way to overcome this issue is to assemble a whole genome, or construct a linkage map for the species being studied, so that an accurate assessment of LD can be taken into account before using methods such as RADseq (Lowry *et al.* 2017b). However, assembly of the Atlantic salmon (*Salmo salar*) genome took more than six years, and the collaboration of 45 (named) researchers from three countries (Lien *et al.* 2016). Furthermore, the use of incorporating already existing genomic resources into studies using RADseq to detect local adaptation (Chapter 3) can also strengthen the evidence. Generally, my opinion is that concerns over LD and detecting the entirety of the genome underlying local adaptation are valid, but, until such time as whole-genome sequencing becomes significantly cheaper, and assembling genomes becomes computationally more efficient, a 'glass half full' perspective on detecting local adaptation using RADseq should persist.

7.3 HOW FAR DOWN THE EVOLUTIONARY ROAD ARE METAL-TOLERANT TROUT?

Given the genetic distinctiveness of locally adapted brown trout, what does this mean for the evolutionary and taxonomical classification of these populations? And what implications does this have for the conservation and management of this highly variable species? Identifying regions of the genome involved in local adaptation can assist conservation and management decisions for preserving the rich diversity within a species, and also how the performance of such genotypes can operate in novel environments. Metal-tolerant trout constitute an example of one of several fish species whose evolution has been driven by human-altered environments. Such a discovery is important not only from an evolutionary perspective, but also in terms of conservation. As a result of their adaptation, these

populations have important evolutionary significance, and, given the current pace of environmental change, should be viewed as representing populations harbouring unique genetic resources.

Due to their 'adaptive distinctiveness' (Crandall *et al.* 2000) and the results presented in this thesis, each of these metal-adapted populations could be classified as representing evolutionary-significant units (ESUs), or at the very least, as management units (MUs) Importantly, ESUs are both demographically and evolutionary distinct, whereas MUs can be defined as demographically independent, or isolated, but not necessarily evolutionary distinct (Palsbøll *et al.* 2007). Further research addressing the phylogenetic history of these metal-adapted populations would assist the designation of these populations as ESUs or MUs. Nonetheless, management of these metal-adapted ESUs/MUs should seek to preserve their unique genetic distinctiveness. Whilst supplemental stocking of these rivers with foreign populations may seem desirable, given the evidence of population bottlenecks (Chapter 2) and low effective population sizes (Chapter 4), such stocking risks swamping these unique genetic assemblages. Furthermore, as we have shown (Chapter 5 and Chapter 6), metal-naïve fish would likely not persist in these metal-rich rivers (Bury *pers. comm.*).

Through successive demographic changes, including historical and more recent bottlenecks, as well as selective processes, these populations now contain a reduced genetic repertoire. Such reduced levels of genetic diversity and the small population sizes observed in these trout populations are concerning and will likely hinder genetic buffering of the populations from future environmental change and local stochastic events (Lane & Shannon 1996; Hughes *et al.* 2008; Reed *et al.* 2003; Frankham *et al.* 2014). The correlation between fitness and genetic diversity has been a major focus of population genetics for a long time (Reed & Frankham 2003), and it has been evidenced that populations with reduced diversity show a lack of resilience to other environmental stressors; for example, in experimental populations of *Drosophila*, inbreeding and decreased genetic variation reduced resistance to the stress of disease (Frankham 2005). Thus, 'evolution may not be the

solution to pollution' (Whitehead et al. 2017) in regards to the long-term viability of these adapted but genetically depauperate brown trout populations.

7.4 LIMITATIONS AND AVENUES FOR FUTURE RESEARCH

There are several limitations to these investigations, the drawbacks of which actually represent exciting new avenues for research into exploring further these metal-adapted trout populations. Perhaps most interesting from an evolutionary perspective is quantifying the fitness costs of metal adaptation. Although the laboratory-based experiment conducted here (Chapter 5 and Chapter 6) elucidated several mechanisms of metal physiology, a reciprocal transplant experiment would not only expose metal-naïve fish to environmentally realistic metal concentrations, but would provide an experimental basis for exploring the costs of metal adaptation in metal-tolerant fish. Such costs could be measured in the form of health indices measured in 'clean' water, the ability to become anadromous (can metal adapted fish smolt and go to sea?) and energy expenditure in terms of growth and reproduction.

As mentioned in the Introduction, the processes of local adaptation explored here were limited to primarily investigating microevolutionary processes. Metal adaptation in these fish could also be facilitated by genomic architectures not interrogated here. In particular, gene duplications may play a role in conferring resistance to metals, especially given that salmonids are pseudotetraploids. In the publication associated with the release of the Atlantic salmon genome, the authors highlight the role that genome duplication has likely played in the ecological flexibility of salmonids (Lien *et al.* 2016). A sequenced genome of the brown trout would certainly assist such explorations. In attempts to quantify specific regions of the genome underlying selection, we used the phylogenetically nearest reference genome of the Atlantic salmon. Although brown trout (*Salmo trutta*) and Atlantic salmon (*Salmo salar*) represent two distinct species within the same genus, and gene synteny is conserved between the species (Gharbi *et al* 2006), significant genomic divergence is expected between the two species, and so we admit to the shortfalls of using the salmon genome for exploring islands of genomic divergence in brown

trout. Indeed, alignment to a “true” reference may result in observances of highly conserved regions of the genome underlying selection, as observed in studies of freshwater adaptation in three-spine stickleback (Roseti *et al.* 2012; Jones *et al.* 2012; Catchen *et al.* unpublished). Furthermore, Atlantic salmon typically have 29 chromosomes, whereas brown trout have far more (mean of 40). Recently, an improved linkage map has been developed for the brown trout, which showed that in comparison to the Atlantic salmon, brown trout show extensive chromosomal rearrangements (Leitwein *et al.* 2017). This genetic resource should be leveraged in future genomic investigations regarding metal-adapted brown trout populations. Further, the role of epigenetics in conferring metal tolerance should be explored; this has shown to be an important component of metal tolerance in other fish species, for example, in copper exposure in stickleback (Laing *et al* unpublished).

Lastly, the scale and intensity of metal mining across the southwest of England implies that the majority of rivers in the region would have been affected (Pirrie *et al.* 2002, 2003; Thorndycraft *et al.* 2004). Future studies should explore how distinctive the southwest England brown trout populations are as a whole group, especially in comparison to brown trout populations that have not experienced a metaliferous past (for example, fish from the chalk streams of southern England). Furthermore, the metal-impacted populations chosen here represent just a handful of rivers impacted by the mining industry of Britain, but have served to highlight how metal-tolerance can take various forms, depending on the mixture of metals present in a given river. Future research should aim to assess alternative patterns of local adaptation in other populations of trout residing in metal-rich rivers in the region, as well as elsewhere in Britain, and further afield.

In conclusion, an ideal assessment of local adaptation will unpick the underlying physiological processes that drive the patterns of local adaptation elucidated using genetic and genomic techniques. Both an empirical and theoretical understanding of these processes together are essential in verifying local adaptation, and in demonstrating the resilience of a species and its capacity to adapt to human-altered environments.

REFERENCES

Catchen, J. M., Hohenlohe, P. A., Bernatchez, L., Funk, W. C., Andrews, K. R., & Allendorf, F. W. (2017). Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. *Molecular Ecology Resources*. DOI: 10.1111/1755-0998.12669.

Chevin, L.-M., Lande, R., Mace, G. M., Sheldon, K. S., & Ghalambor, C. K. (2010). Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory. *PLoS Biology*, **8**(4), e1000357.

Courtenay, S. C., Grunwald, C. M., Kreamer, G.-L., Fairchild, W. L., Arsenault, J. T., Ikonomou, M., & Wirgin, I. I. (1999). A comparison of the dose and time response of CYP1A1 mRNA induction in chemically treated Atlantic tomcod from two populations. *Aquatic Toxicology*, **47**(1), 43–69.

Crandall, K. A., Bininda-Emonds, O. R. P., Mace, G. M., & Wayne, R. K. (2000). Considering evolutionary processes in conservation biology. *Trends in Ecology & Evolution*, **15**(7), 290-295.

Durrant, C. J., Stevens, J. R., Hogstrand, C., & Bury, N. R. (2011). The effect of metal pollution on the population genetic structure of brown trout (*Salmo trutta* L.) residing in the River Hayle, Cornwall, UK. *Environmental Pollution,* **159**(12), 3595–603.

Fernandez, M. P., Ikonomou, M. G., Courtenay, S. C., & Wirgin, I. I. (2004). Spatial variation in hepatic levels and patterns of PCBs and PCDD/Fs among young-of-the-year and adult Atlantic tomcod (*Microgadus tomcod*) in the Hudson River estuary. *Environmental Science & Technology*, **38**(4), 976–83.

Finn, R. N. (2007). The physiology and toxicology of salmonid eggs and larvae in relation to water quality criteria. *Aquatic Toxicology*, **81**(4), 337–354.

Frankham, R. (2005). Stress and adaptation in conservation genetics. *Journal of Evolutionary Biology*, **18**(4), 750–755.

Frankham, R., Bradshaw, C. J. A., & Brook, B. W. (2014). Genetics in conservation management: Revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. *Biological Conservation*, **170**, 56–63.

Gharbi, K., Gautier, A., Danzmann, R. G., Gharbi, S., Sakamoto, T., Høyheim, B., *et al.* (2006). A linkage map for brown trout (*Salmo trutta*): chromosome homeologies and comparative genome organization with other salmonid fish. *Genetics*, **172**(4), 2405–19.

Hamilton, S. J., & Mehrle, P. M. (1986). Metallothionein in fish: review of its importance in assessing stress from metal contaminants. *Transactions of the American Fisheries Society*, **115**(4), 596–609.

Hughes, A. R., Inouye, B. D., Johnson, M. T. J., Underwood, N., & Vellend, M. (2008). Ecological consequences of genetic diversity. *Ecology Letters*, **11**(6), 609–623.

Jeffries, D. L., Copp, G. H., Lawson Handley, L., Olsén, K. H., Sayer, C. D., & Hänfling, B. (2016). Comparing RADseq and microsatellites to infer complex phylogeographic patterns, an empirical perspective in the Crucian carp, *Carassius carassius,* L. *Molecular Ecology*, **25**(13), 2997–3018.

Jones, F. C., Chan, Y. F., Schmutz, J., Grimwood, J., Brady, S. D., Southwick, A. M., Absher, D.M., Myers, R.M., Reimchen, T.E., Deagle, B.E., Schluter, D., & Kingsley, D. M. (2012). A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. *Current Biology*, **22**(1), 83-90.

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, **7**(12), 1225–1241.

Lande, R., & Shannon, S. (1996). The role of genetic variation in adaptation and population persistence in a changing environment. *Evolution*, **50**(1), 434.

Laurén, D. J., & McDonald, D. G. (1987). Acclimation to copper by rainbow trout, *Salmo gairdneri*: physiology. *Canadian Journal of Fisheries and Aquatic Sciences*, **44**(1), 99–104.

Leitwein, M., Guinand, B., Pouzadoux, J., Desmarais, E., Berrebi, P., & Gagnaire, P.-A. (2017). A dense brown trout (*Salmo trutta*) linkage map reveals recent chromosomal rearrangements in the *Salmo* genus and the impact of selection on linked neutral diversity. *G3,* **7**(4), 1365–1376.

Lien, S., Koop, B. F., Sandve, S. R., Miller, J. R., Kent, M. P., Nome, T., *et al.* (2016). The Atlantic salmon genome provides insights into rediploidization. *Nature*, **533**(7602), 200–205.

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer, A. (2017a). Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Molecular Ecology Resources*, **17**(2), 142–152.

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer, A. (2017b). Responsible RAD: Striving for best practices in population genomic studies of adaptation. *Molecular Ecology Resources*. DOI: 10.1111/1755-0998.12677.

McKinney, G. J., Larson, W. A., Seeb, L. W., & Seeb, J. E. (2017). RADseq provides unprecedented insights into molecular ecology and evolutionary genetics: comment on Breaking RAD by Lowry *et al*. (2017). *Molecular Ecology Resources*. DOI: 10.1111/1755-0998.12649.

Palsbøll, P., Berube, M., & Allendorf, F. (2007). Identification of management units using population genetic data. *Trends in Ecology & Evolution*, **22**(1), 11–16.

Pirrie, D., Power, M.R., Rollinson, G., Camm, G.S., Hughes, S.H., Butcher, A.R., & Hughes, P. (2003). The spatial distribution and source of arsenic, copper, tin and zinc within the surface sediments of the Fal Estuary, Cornwall, UK. *Sedimentology*, **50**, 579-595.

Pirrie, D., Power, M.R., Wheeler, P.D., Cundy, A., Bridges, C., & Davey, G. (2002). Geochemical signature of historical mining: Fowey Estuary, Cornwall, UK. *Journal of Geochemical Exploration*, **76**, 31-43.

Radwan, J., & Babik, W. (2012). The genomics of adaptation. *Proceedings of the Royal Society of London B: Biological Sciences*, **279**(1749), 5024-8.

Reed, D. H., & Frankham, R. (2003). Correlation between fitness and genetic diversity. *Conservation Biology*, **17**(1), 230–237.

Reid, N. M., Proestou, D. A., Clark, B. W., Warren, W. C., Colbourne, J. K., Shaw, J. R., *et al.* (2016). The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. *Science*, **354**(6317), 1305-1308.

Roesti, M., Hendry, A. P., Salzburger, W., & Berner, D. (2012). Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. *Molecular Ecology*, **21**(12), 2852–2862.

Roy, N. K., Courtenay, S., Maxwell, G., Yuan, Z., Chambers, R. C., & Wirgin, I. (2002). Cytochrome P4501A1 is induced by PCB 77 and benzo[a]pyrene treatment but not by exposure to the Hudson River environment in Atlantic tomcod (*Microgadus tomcod*) post-yolk sac larvae. *Biomarkers*, **7**(2), 162–173.

Stapley, J., Reger, J., Feulner, P. G. D., Smadja, C., Galindo, J., Ekblom, R., *et al.* (2010). Adaptation genomics: the next generation. *Trends in Ecology & Evolution*, **25**(12), 705–712.

Thorndycraft, V. R., Pirrie, D., & Brown, A. G. (2004). Alluvial records of medieval and prehistoric tin mining on Dartmoor, southwest England. *Geoarchaeology*, **19**(3), 219–236.

Tonteri, A., Vasemägi, A., Lumme, J., & Primmer, C. R. (2008). Use of differential expression data for identification of novel immune relevant expressed sequence tag-linked microsatellite markers in Atlantic salmon (*Salmo salar* L.). *Molecular Ecology Resources*, **8**(6), 1486–1490.

Uren Webster, T. M., Bury, N., van Aerle, R., & Santos, E. M. (2013). Global transcriptome profiling reveals molecular mechanisms of metal tolerance in a chronically exposed wild population of brown trout. *Environmental Science & Technology*, 47(15), 8869–77.

Whitehead, A., Clark, B. W., Reid, N. M., Hahn, M. E., & Nacci, D. (2017). When evolution is the solution to pollution: Key principles, and lessons from rapid repeated adaptation of killifish (*Fundulus heteroclitus*) populations. *Evolutionary Applications*. DOI: 10.1111/eva.12470.

Wirgin I.I. & Chambers R.C. (2006). Atlantic Tomcod *Microgadus tomcod*: A model species for the responses of Hudson River fish to toxicants. In: J. Waldman, K. Limburg, and D. Strayer (eds.) *Hudson River fishes and their environment*, American Fisheries Society Symposium, **51**, pp. 331-365.

Wirgin, I. I., Grunwald, C., Courtenay, S., Kreamer, G. L., Reichert, W. L., & Stein, J. E. (1994). A biomarker approach to assessing xenobiotic exposure in Atlantic tomcod from the North American Atlantic coast. *Environmental Health Perspectives*, **102**(9), 764–70.

Wirgin, I. I., Kreamer, G.-L., Grunwald, C., Squibb, K., Garte, S. J., & Courtenay, S. (1992). Effects of prior exposure history on cytochrome P4501A mRNA induction by PCB congener 77 in Atlantic Tomcod. *Marine Environmental Research*, **34**(1-4), 103–108.

Wirgin, I., Roy, N. K., Loftus, M., Chambers, R. C., Franks, D. G., & Hahn, M. E. (2011). Mechanistic basis of resistance to PCBs in Atlantic Tomcod from the Hudson River. *Science*, **331**(6022), 1322-1325.