Population Structuring of the Freshwater Fish, Roach (*Rutilus rutilus*, L. 1758): Implications for Migration and Management

Steven Crookes

Thesis submitted to the University of London for the degree of Doctor of Philosophy

Evolution, Ecology and Environment Group, School of Biological Sciences, Royal Holloway, University of London, 2013

Declaration of Authorship

I, Steve Crookes, hereby declare that this thesis and its written contents are entirely of my own creation. Where the work of others has been incorporated into the text, the proper citation has been accredited and all authors and contributors of data, quotes and maps have been fully recognized.

Signed:

Date:

Acknowledgements

Gratitude is extended to my supervisor, Prof. Paul Shaw, for advice on the thesis and professional support during a difficult period in my life. I would also like to thank the following for providing samples: Katy Lewis of the Environment Agency, UK, for providing all UK samples of cycloid scales; Lene Jacobsen (Danish samples), Henri Persat (French samples), Markku Viljanen and Prof. Jouko Sarvala (Finnish samples), Dirk Neumann (German samples) and Serik Timirkhanov (Capsian roach, Caspian Sea) for the provision of European roach scales. I also wish to thank Richard Williams and Nicky Tombs, of the Centre for Ecology and Hydrology, Wallingford, for providing oestradiol and hydrometric GIS data, respectively. Finally, my thanks to Nigel Hewlett, Stephen Axford and Graeme Peirson of the Environment Agency for providing pertinent information regarding the coarse fishery of England and Wales.

Dedicated to mum. This is for you.

Abstract

The roach is a key species component of the economically and culturally significant recreational angling community. It is also a keystone species by which the health of a coarse fishery can be judged. However, despite being commonplace and widely distributed in the British Isles, very little is known about the underlying genetic structuring of this species, which could inform a more inclusive, synergistic management system, thus improving an already billion pound strong industry. This study describes and explains the extent of genetic structuring at hierarchical levels within the putative natural distribution of this species within the British Isles, and to uncover the causes, both historical and contemporaneous, for any non-homogeneous distribution of genetic diversity. Mitochondrial DNA was utilised to account for coarse-grained structuring across catchments to divulge whether any individual catchment or sets of drainages could be considered as individual management units. Low levels of sequence diversity were observed across the UK, consistent with a history of recent post-glacial demographic expansion, although haplotype assemblages within and among river populations differed. The inclusion of microsatellite data allowed a higher resolution analysis of structuring between and within sampled sub-populations of two southeastern rivers: the Thames and the Suffolk Stour. Despite sharing high levels of within-population genetic diversity, significant differentiation was observed among populations within both rivers. Isolation by distance is observed in both rivers, indicating genetic equilibrium had been attained among populations from which estimates of recent migration were obtained. Roach in the Stour embody a source-sink model of equilibrium, explaining its higher levels of inter-population differentiation and rates of migration than the Thames population. The discovery of significant genetic differentiation suggests that one must apply caution in managing this diversity. Stocking from exogenous sources is not recommended unless there exists no other choice.

Contents

Chapter Four – Within-Population Microsatellite DNA Diversity and Among-Population Differentiation of Roach Within and Between Two Historically Connected Rivers in the Southeast of England 147

List of Tables

List of Figures

Chapter One - General Introduction

"*The roach is an easy fish to catch. And if he is fat and penned up, then he is good food, and these are his baits. In March, the readiest bait is the red worm. In April, the grub under the cow turd. In May, the bait that grows on the oak leaf and the grub in the dunghill. In June, the bait that grows on the osier and the codworm. In July, houseflies and the bait that grows on all oak; and the nutworm and mathewes and maggots till Michaelmas. And after that, the fat of bacon*." **Dame Juliana Berners – A Treatyse on Fisshnge with an Angle (1496) (the first recorded historical account, and manual, of coarse fishing).**

1.1 Coarse Fisheries in England and Wales

The coarse fishery of England and Wales represents a significant economic component of the overall inland fishery of these enthusiastic angling nations. Chief amongst the patrons of some 26000km of linear waterways, which are utilized for leisure purposes (with some 40000km unused, but potentially open for exploitation), are recreational anglers (Peirson et al. 2001). Recreational fishing, whereby fish are caught primarily as a leisure activity (Pitcher and Hollingworth 2002), is a familiar and highly popular pastime in the UK and throughout the world (Cowx 2002). The vast majority of anglers catch fish for sport, focussing on species categorised as 'coarse', including the cyprinids, perch *Perca fluviatilis*, pike *Esox lucius* and the European eel *Anguilla Anguilla* (Pierson et al. 2001). Unlike the game fish (e.g. salmonids), where there exists an element of fishing for personal consumption (Pitcher and Hollingworth 2002), coarse fish are usually returned to the water without lasting impact upon their numbers in the long term (Wortley 1995).

In 1994, there were some 2.4 million coarse anglers that collectively had an economic impact - through the purchase of rod licenses, tackle, bait, travel costs and other angling-related paraphernalia - in excess of three thousand million pounds (Moon and Souter 1994). A study by Robinson et al. (2003) using catch data from angling competitions further underlines the economic importance of the angling community. As of 2003, over 2 million anglers each spent, on average, approximately £1000 over the course of a year based upon an average of 43 trips per annum. In 2008, record numbers of rod licenses

were sold by the Environment Agency (1.3 million), such that the contribution in rod license fees alone to the angling economy had risen to £1 billion per year in the intervening 14 years (http://www.environment-agency.gov.uk/news/104693.aspx). In addition to the fiduciary benefits to the local and national economies, inland fisheries, including the coarse fishery, exert an important influence upon the social fabric of their local communities (Weithman 1999).

Freshwater fisheries, as has been documented, provide an important economic boon to the UK. Befitting an important source of income, coarse fish have been protected under the rubric of parliamentary acts since the 1970s (e.g. the Salmon and Freshwater Fisheries Act 1975 and the Environment Act of 1995 (Hickley 1995)) when the rapid deterioration of riverine habitats, particularly from industrial and domestic pollution, became a major concern. Originally under the auspices of the National Rivers Authority (NRA), but lately under the purview of the Environment Agency (EA), the UK's rivers are afforded significant protection both at the preventative and at the prohibitive (legislative) level. However, the endeavours of the EA are hampered by the fact that the majority of the waterways under its influence are adjacent to private lands over which the EA has little or no control. Thus, pollution events are unlikely to be completely eradicated and, as a result of such episodes, recuperative management protocols are necessary for the rehabilitation of significant stretches of rivers after large fish kills.

The EA, in its role as the chief guardian of the national stock of freshwater fish and all fluvial habitats (and all recreational fishing (Lyons et al. 2002)), has a dual responsibility: to maintain, conserve and develop fisheries; and to allow access to every citizen of England and Wales the right to 'experience a diverse range of good quality fishing' (Environment Agency 1999). To fulfill the first part of this charter, the EA is proactive in maintaining and improving fish numbers, species composition, habitat quality and the facilitation of fish passage upstream via fish passes, etc. These management measures are known as Fisheries Action Plans (FAPs) and are generally specifically tailored to the needs of particular catchments (Robinson and Whitton 2004). Much of this work involves consultation with the public and with angling groups, which may lead to conflicts of interest between the EA's dual responsibilities (Hickley and Chare 2004). The impact of anglers (specifically their demand for productive river systems) upon fish stocks has been given scant attention in governance or the fisheries literature (Cooke and Cowx 2006).

Approximately 4500 distinct introductions of fish (*in toto* 1.5 million individuals) are conducted annually in England and Wales (Hickley and Chare 2004) on a remit of mitigation, restoration, enhancement or the genesis of new fisheries (Hickley 1994). A significant proportion of these introductions are to maintain recreational fisheries (Harris 1978; Maitland 1986; and Cowx 2002). Of particular concern, due to its widespread nature, is the potential for the mismanaged enhancement of stock through the introduction of allochthonous fish, which may have serious detrimental effects upon the resident fish biota (Cowx 1998). Reasons for the failure of a stocking event are myriad, but may be dichotomized into genetic and non-genetic effects. Non-genetic effects include the competitive disadvantage of introduced fish and increased susceptibility to predation; also, introduced fish may be susceptible to parasites and disease already *in situ*; or may themselves bring with them parasites which may have a deleterious impact upon the endogenous populations (see Cowx & Gerdeaux 2004 and Cooke & Cowx 2006 and references therein). Genetic effects of introductions involve the erosion of the genetic variation of endogenous stock (Millar and Libby 1994), the breaking up of localized adaptive genic complexes via crossbreeding, resulting in so-called outbreeding depression (Templeton 1986), and the general decrease of genetic diversity as a whole across the species' distribution (Maitland 1979). However, the influx of 'new' genetic variation via crossbreeding with exogenous stock may revitalize locally inbred populations which otherwise would experience reduced fitness (Hodder and Bullock 1997). A full exegesis of the theory of the evolutionary and conservation genetics of freshwater fish populations and the effects of stocking are discussed on pages 30-37 and in the ensuing chapters. However, stocking, for the recuperation of a fishery, should only be conducted in areas in which natural recruitment from local spawning and nursery areas are insufficient to meet the demands of the local angling community and recuperation stratagems (Maitland 1987; Hodder and Bullock 1997). Additionally, much care needs to be exercised in stocking depauperate areas such that they do not become the fluvial equivalent of agricultural monocultures.

Despite their potentially negative influence, anglers are important in helping to fund the continued maintenance and rehabilitation of the UK's waterways. They may also provide useful data and insight, via their preferences, into what constitutes a healthy coarse fishery. A quantitative measure of fisheries production is the CPUE (catch per unit effort), a metric that has been exapted from commercial fisheries to determine the mass of fish caught per unit time per capita (angler). Although in angling circles the CPUE may be biased by the skills of particular anglers over others, amongst the most productive of the coarse fish is

the common roach *Rutilus rutilus* (L. 1758). The presence of roach, among other smaller species of freshwater cyprinids, in rivers yields greater CPUE per angler than where such fish are lacking (Robinson et al. 2003). High levels of CPUE translate into higher levels of angler satisfaction, having a positive knock-on effect on the coarse fishery economy, as angler satisfaction correlates with the perceived health of a fishery. Also, whilst the popularity of the roach as preferred catch has waned slightly in the last 30 years or so (39% of anglers surveyed by the NRA fell to 28% from 1979 – 1995 (National Rivers Association 1995)), due in part to the rise in the introductions of exotic species (Hickley and Chare 2004), the continued persistence of roach, as evidenced by high catch rates (45% of all fish caught in 1994 were roach (National Rivers Association 1995)), suggests that the roach may be useful as an indicator species for the general health of a productive coarse fishery. Thus, it is in the interests of the EA and private fisheries managers to maintain healthy stocks of the roach and other coarse fish.

1.2 The Roach: A Biography.

1.2.1 Brief Systematic and Biogeographic Review

Rutilus rutilus (or the 'common' or 'minnow' roach (but hereafter referred to as 'the roach')) is a member of the genus of 'roaches', a moderate, mostly peri-Mediterranean distributed genus belonging to the minnow sub-family (Leucisinae) of the Cyprinidae[1]. Although the taxonomy of the cyprinids has elicited controversy due to difficulties in elucidating genealogical relationships based upon morphological characters, recent molecular work (Gilles et al. 1998, 2001; He et al. 2008; Luca et al. 2008) reaffirms the hierarchical taxonomy listed above. Close genealogical ties with the leuciscine fishes[2], such as members of the genera *Abrama*, *Scardinius* and *Leuciscus*, is reinforced by the fact that the roach naturally hybridises with species in each of these genera (Kennedy and Fitzmaurice 1973; Wheeler 1976; Wheeler and Easton 1977; Cowx 1983; and Adams and Maitland 1991).

[1]The cyprinids (Class: Actinopterygii; Order: Cypriniformes) encompass the minnow and carp species and is also the fish family in possession of the greatest number of freshwater species (it also competes with the teleost Gobiidae for the title of largest family of vertebrates). The Cyprinidae consist of some 2420 species in over 220 genera (Nelson 2006).

[2] Leuciscinae is the largest of the cyprinid sub-families found in Europe (Kottelat and Freyhoff 2007).

Rutilus encompasses some 17 known species (www.fishbase.org), further taxonomic splitting of subspecies and cryptic 'varieties' notwithstanding. The roaches vary in size but are mostly small or medium in extent. Most of the genera are found in low numbers and limited distributions within the peri-Mediterranean region of Europe (an area rich in endemic freshwater fish (Smith and Darwell 2006)), and vary greatly in their ecology, behaviour and natural history. Although low on species diversity,, roaches vary greatly in their morphology, buttressed by molecular evidence suggesting a middle Miocene splitting of the lineages that lead to the morphologically divergent common roach and the Danube roach *Rutilus pigus* (Ketmaier et al. 2008). *R. rutilus* is the most successful and widespread of all the roaches. Whereas the other species have very limited distributions, the roach is found throughout western and northern Europe (up to 69^{o}N in Scandinavia and 56^{o}N in Scotland), southern Europe to the north of the Pyrénées, and are found as far eastwards and southwards as the Ural mountains, the Balkans and the catchments that drain into the Aegean, Black and Caspian seas (Kottelat and Freyhof 2007). In Britain, the natural distribution of the roach includes mostly those rivers that flow into the North Sea, the English Channel and the southern Irish Sea, with higher densities of fish in the southeast and central England (Carter 2004**)** (see Fig 1.1).

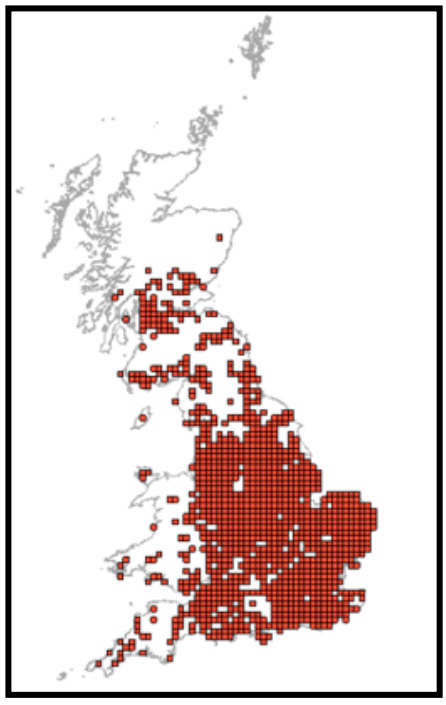


Figure 1.1: Distribution of roach in Great Britain. Each red square represents the occurrence of roach within an area of 10km^2. The data is from the Database and Atlas of Freshwater Fishes (DAFF) project, initiated by PS Maitland in 1966, collated by the Biological Records Centre (BRC) and held online at the National Biodiversity Gateway Network (NBN, http://data.nbn.org.uk); and includes both native and introduced roach.

1.2.2 Roach Morphology and Behaviour

The roach possesses a laterally compressed fusiform shape adaptive for rapid movements, enabling the roach to capture food and evade predators over short distances. The roach's scientific name references the ruby colouration of its anal and pectoral fins and the irises of its eyes (see Fig 1.2). Elsewhere, its colouration reflects semi-cryptic adaptation to life in rivers: its dorsal body is covered by dark green/blue scales to limit detection from above; conversely its flanks are covered by silvery cycloid scales which may aid in predator avoidance by scattering light and thus diminishing its image in the eyes of nearby predators in the lower-middle water column. Ventrally, the roach is pale.

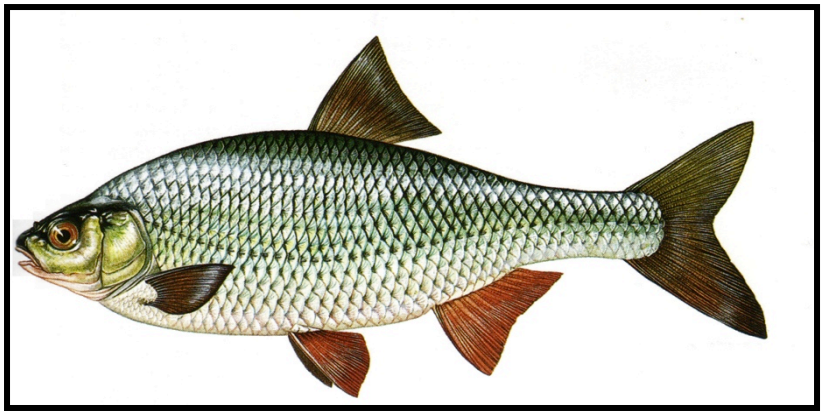


Figure 1.2: Illustration of an adult roach revealing the detailed colouring and morphology of a typical individual. Illustration by Keith Linsell (taken from Maitland and Linsell 2006).

However, roach may display variation in morphology and behavior, depending upon ecological circumstance. The juvenile roach of the Rybinskoe Water reservoir, Russia, show morphological divergence in juvenility according to favoured prey types and the ecosystem in which the prey is found: demersal roach, feeding upon shellfish; and a coastal, non-specific feeding ecotype (Stolbunov and Gerasimov 2008). They differ morphologically in the size and shape of the mouth (the demersal form possessing larger mouths), and in the shape of the body. Differentiation into the two ecotypes occurs early in development, which may indicate a non-plastic element to their juvenile ontogeny (Stolbunov and Pavlov 2005). Whilst riverine and lacustrine roach mostly inhabit slow moving streams and lake habitats of inland areas, as with the Russian roach, they have been known to be found in estuarine environments in locations across Europe (e.g. River Odra, Poland (Więsk and Załakowski 2000), the Belgian Scheldt (Maes et al. 2005) and in the brackish waters of the Baltic Sea (e.g. off the Finnish coast (2-6% salinity levels, Wilklund et al. 1996)). Additionally, they are found in the tidal stretches of most rivers

and can tolerate, as an upper boundary, 0.34 – 2.96 practical salinity units (psu) in the Thames. However, the roach in the upper Thames estuary are recruited from upstream when flows are high and salinity levels drop (from August – November) (Araújo et al. 1999). Generally, salinities higher than 10% are fatal to freshwater fish (Pethon 1980). Whether the estuarine/tidal roach are selectively differentiated in their morphology, behavior or osmoregulatory mechanism from conspecifics elsewhere in the potamodromous environment is unknown.

Roach are gregarious at all stages of life, thereby increasing an individual's chances of surviving and locating food through foraging efficiency (Beecham and Farnsworth 1999). A well-developed lateral line system is evidenced in roach from an early age. Yearling roach (0+) possess an innate ability to elicit a scatter-predator avoidance reaction in response to infrasound waves generated by predators (Karlsen et al. 2004). Ontogenetic shifts in habitat preference during growth also reflect not only the availability of food but also a trade-off with visibility to predators, especially piscivorous predators, when young (Copp 1990; Garner 1996). Larval and juvenile roach feed amongst the benthos in littoral areas which also provide protection from predation and from high flows (Mann 1996), initially preferring deeper areas up to a metre deep with thick macrophyte densities before moving to the shallows (Copp 1992). Diel movements begin when juvenile: young roach in the Thames aggregate at night, more so than in diurnal hours (Matthews 1971), probably in response to predation by the European perch and the chub *Squalius cephalus* (Copp & Jurajda 1993). As they mature, roach start to take advantage of the plentiful supply of the pelagic zooplankton, possibly as a means of increasing growth rates to elevate them from the size range preferred by the piscivorious perch– a major predator and with whom they are competitive with for food when adult. In the first year of life roach can grow at a rate of 30% of the body weight per individual per day, such that young-of-the-year fish attain lengths of between 5-7cm after twelve months of growth (Everard 2006).

When roach attain sexual maturation[3] they reach lengths of 20-40cm (Carter 2004), but large males may exceed this range (e.g. 52cm (Maitland and Linsell 2006)). Adults tend to return to feed in the littoral zone feasting upon detritus and benthic invertebrates. The roach is a mobile species, employing favorable grounds in which to feed (this can be pelagic and/or littoral, depending upon life-history stage, presence of predators and/or

[3] Sexually mature individuals – the roach's somatic ontogeny is, like many other fish, indeterminate, such that a fish keeps on attaining ever-greater size (dependent upon availability of food) until death.

competition, and upon time of the day). Roach, like many other freshwater fishes, show greatest levels of activity at low light, especially the hours around dusk and dawn (Lucas and Baras 2001). Roach migrate to over-winter in slow, deep backwaters and lacustrine environments (Copp 1997; Jepsen and Berg 2002; and Kottelat and Freyhof 2007), to avoid predation when food resources are low and associated risks are high; and to spawn (Goldspink 1977; Diamond 1985; and L'Abée-Lund & Vøllestad 1985). The roach makes use of limited home ranges, (~3km average, although sometimes long distance foraging takes place over stretches of river up to 10km) (Baade & Fredrich 1998)) that it uses in the summer months to replenish the energy lost during spawning and in preparation for the exertions of the subsequent spawning season the following year. A female may lose, as expelled ova, as much as 15% of her body weight during spawning (Everard 2006). Thus the ability to freely move within both lotic and lentic habitats is an important component in the ecology of this species.

1.2.3 Reproductive Biology

1.2.3.1 Spawning Migrations

The roach is an iteroparous species, reproducing annually after sexual maturation until death. Males reach reproductive maturity at around 3 years of age, and females a year after. Spawning migration in roach is comprised of two components: a pre-spawning migration and a final migration to the actual spawning grounds (Everard 2006). This en masse migration to reproduce is believed to be associated with hormonal reactions, associated with the final maturation of gonads and gamete production, to photoperiod (Jafri 1987) and to water temperature (Gillet and Quétin 2006). However, the increasing average water temperatures of the preceding decades has altered the initiation of spawning migrations, indicating that water temperature is the more important determinant (Nõges and Järvet 2005; Gillet and Quétin 2006; and Härmä et al. 2008). Climate alteration notwithstanding**,** roach traditionally begin migration from April through June, when water temperatures are between 10^{o}c and 15^{o}c (Vøllestad & L'Abée-Lund 1987; Gillet and Dubois 1995; Prignon et al. 1998; and Travade et al. 1998).

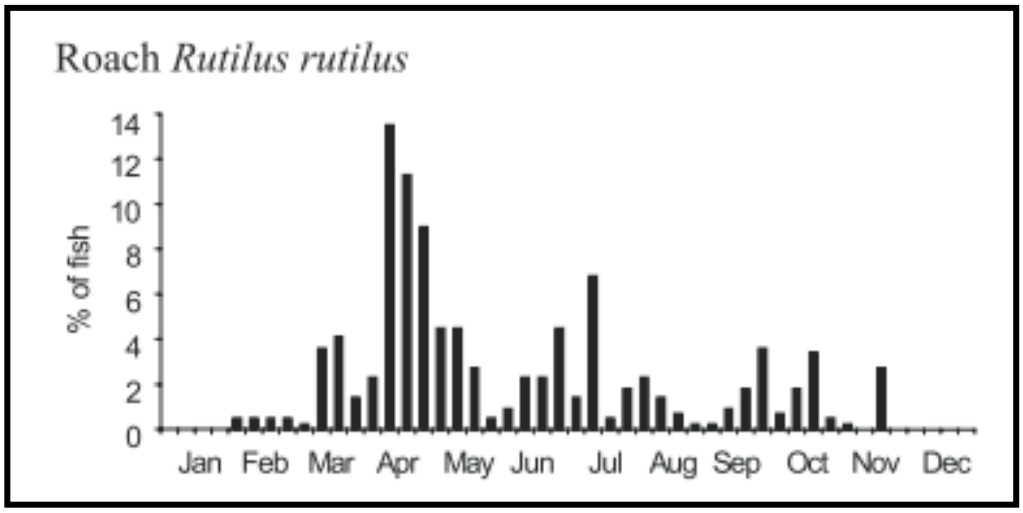


Figure 1.3: Roach migratory activity. Data collated from numerous European rivers (see text for details; taken from Lucas and Baras 2001).

Fig 1.3 summarizes the migration activity of roach, counted as the number of fish migrating upstream through fish passes, over 'one year' (a summary of a number of years' worth of data). The data are collated from a number of western European rivers, including the Garonne and Dordogne, France (Travade et al. 1998), the Meuse and Mehaigne in Belgium (Prignon et al. 1998; Philippart et al. unpublished), the Dutch stretch of the Meuse (Lanters 1993, 1995) and the Mosel, Germany (Pelz 1985). Most movement through the fish passes occurs during April and May, consonant with the spawning period (pre-empted by a pre-spawning migration in March). The peak in movement in late June and early July is probably reflective of post-spawning migration back to the feeding grounds. The Autumnal peaks show passage to overwintering areas, while in December and January there is no fish movement at all.

At least in lacustrine environments roach are known to exhibit strong tendencies to migrate to annual spawning grounds. 83.5 – 92.0% of all roach studied returned to the same spawning ground in the Norwegian Lake Årungen (L'Abée-Lund & Vøllestad 1985). Whilst these figures are comparable with the better-known homing abilities of salmonids, it is unknown whether the roach returned to their own natal spawning area (Lucas and Baras 2001). Diamond's study (1985) suggests that roach in both lotic and lentic environments utilized the same spawning grounds year after year. The observation of spawning homing abilities in roach is further evidenced by the finding that reciprocally translocated individuals from two spawning grounds 3km apart migrated back to the spawning ground from which they were taken (Goldspink 1977).

Spawning migrations in roach, as in most potamodromous fish, are usually upstream (Lucas and Baras 2001). In lacustrine environments suitable spawning substrate is only to be found in feeder tributaries, e.g. up the River Endrick, which feeds Loch Lomond, in Scotland (Adams 1994). However, in the River Axe, Devon, UK, (between 1960-1969 at least), large movements of roach (assumed to be linked to spawning as they occurred during what is the normal spawning season) were in the downstream direction (Champion & Swain 1974). The requirement for spawning substrate in roach is phytolithophilic (Mann 1996); that is they utilize gravel beds, submerged logs and mosses (Vøllestad and L'Abée-Lund 1987), and macrophytes (such as *Fontinalis*, *Phragmites* and *Elodea* spp. (Mills 1981; Everard 2006; and Diamond 1985)) on which to deposit ova and milt. Limitation of spawning substrata for the roach and other phytolithophils (including the common and silver bream, rudd and *Leuciscus* spp.) encourages cross-fertilization amongst them, often resulting in increased incidences of hydrids (Cowx 1983; Mann 1996).

1.2.3.2 Spawning Behaviour

Although the spawning melee seems frenetic and random with respect to the choice of mating partner, recent studies have built upon an original observation, albeit of fish in captivity, of Wedekind (1996), that roach may partake in lek-style aggregations in which the females may exercise some choice over which males are granted access to deposit their milt over the ova. Up to three months prior to spawning, male roach of reproductive age develop tough tubercles on top of, and posterior to, the head, which may aid in 'coupling' during the act of reproduction (Everard 2006). The quality of these tubercles has been positively correlated with host resistance to the gill endo-parasite *Rhipidocotyle campanula* (Taskinen & Kortet 2002) and the over-expression of immuno-suppressant androgens associated with sexual ornamentation and attractiveness to females (Kortet et al. 2003). The corollary of female choice between males exhibiting handicaps (which increases susceptibility to papillomatosis during spawning) as a means to display 'good genes' may be facilitated by the quality of the handicapping signal (tubercles) displayed by the males (Kortet et al. 2004). Thus, male reproductive success may be significantly skewed and mating non-random.

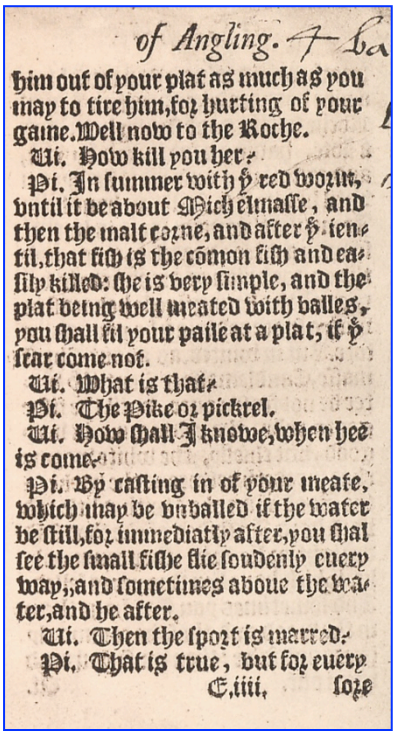
1.3 Roach and Man.

1.3.1 Early History (Pre-History – 19th Century).

Pharyngeal remains characteristic of roach uncovered at Skipsea, Humberside indicate that the roach has been a member of the freshwater piscifauna in the UK since the Devensian glacial period (Carrott et al. 1994). Presumably, by this time, the roach had colonized most of the drainages in the UK in which it is now native. Evidence of fish movement, including roach, date back to the Bronze Age. Roach remains from the River Great Ouse at the Grand Arcade span some 500 years from the 14th – 19th centuries, although most remains were dated prior to the 15th century (Harland 2007). Many of the roach remains (along with dace and rudd) originated from areas associated with medieval fishponds that were commonplace during this period (Aston 1988; Serjeantson and Woolgar 2006 in Harland 2007) and are regarded by historians as fulfilling the role as gifts rather than as a source of food (Dyer 1988). However, the historical account of Berners (1496) explicitly states that if the roach is "fat and penned up, then he makes good food."

1.3.2 Modern Anthropogenic Influence

The post-renaissance histories of roach movements in the UK are tied to the rise of angling – or fly-fishing. Angling as a recreational activity took off in the 15th (Berners 1496) and 16th centuries (Anon. 1577), generating an angling literature that is going strong to this very day (e.g. Everard 2006). Fig 1.4 shows an excerpt from the Arte of Angling (Anon. 1577) on how to catch and kill roach. Although the stocking of local ponds and lakes from nearby rivers has been known since the early period of man's habitation in Britain, mass movements of fish probably only began in earnest with the rapid industrialization of Britain's villages and towns. The resulting pollution of rivers, reaching its nadir during the industrial revolution of the 18th century, severely reduced the biota inhabiting them. However, in the Grand Arcade of the Great Ouse the temporal distribution of roach fossils remained relatively stable compared with those of its cyprinid cousins during the period of the industrialization of nearby Cambridge (Harland 2007), indicating some resistance to the effects of pollution.


of Angling. 4 Ca

him out of your plat as much as you may to tire him, for hurting of your game. Well now to the Roche.

Vi. How kill you her?

Pi. In summer with ỳ red worm, vntil it be about Mich elmasse, and then the malt corne, and after ỳ ientil, that fish is the cõmon fish and easily killed: she is very simple, and the plat being well meated with balles, you shall fil your paile at a plat, if ỳ fear come not.

Vi. What is that?

Pi. The Pike or pickrel.

Vi. How shall I knowe, when hee is come?

Pi. By casting in of your meate, which may be vnballed if the water be still, for immediatly after, you shal see the small fishe flie soudenly euery way, and sometimes aboue the water, and he after.

Vi. Then the sport is marred.

Pi. That is true, but for euery

E.iiii. sore

Figure 1.4: An excerpt from the Arte of Angling, published in 1577, on how to kill roach. The roach is referred to, in old English, as the "Roche" (the French root of our contemporary anglicized word), and the manual is styled as a dialogue between a teacher and pupil, as was the manner of many books of the time.

The roach became, and still is, a highly popular sport fish, becoming highly regarded by sports fisherman the length and breadth of the country. Coarse fishing became increasingly popular with the increase of leisure time available to the lay worker and an increase in tourism advertising (Maitland 1987). As a result, many fish were translocated from native populations into previously virgin territory, especially during the last 200 years. Roach were introduced into Loch Lomond, Scotland by the 1790s (Maitland 1987; Adams 1994). In Ireland, roach were introduced both accidentally (Went 1950) and/or as stocking for angling and baitfish purposes (Hale 1958; Mercer 1968; Kennedy and Fitzmaurice 1973; and Fitzmaurice 1974). Roach were also purposefully translocated to destinations far afield from the UK, e.g. Australia, where the roach was introduced from the UK into the Murray-Darling system, Victoria, in the period 1830-1860 (Arthington and Blühdorn 1995). In each case, the introduction of roach, like numerous other exotic introductions, has generally been to the detriment of native fish species and endogenous ecosystems in general (Fitzmaurice 1984; Cowx 1998; Welcomme 2001; and Barrett 2004).

It is also highly probable, and believed by many authors, that many of Britain's rivers contain non-native populations, especially in the northwest, Wales and Scotland; and that translocations have occurred countrywide between drainages to supplement low productivity of local streams or as baitfish for larger predatory fish (Campbell 1971 in Maitland 1987). The problem remains pertinent as anglers continue to move fish despite the illegality of such translocations (Maitland 1995). Further to the active transport of fish by man, the industrial revolution saw the construction of arterial canal networks that interconnect some of the largest – and most important (at least in the 18th and 19th centuries) - shipping rivers. Such conduits may have facilitated the exchange of freshwater fish along their lines of communication. However, due to the vagility of some species, such as roach, it remains a hard task to determine how much of the national distribution of coarse fish is due to natural or anthropogenic agency (Maitland 1995).

1.4 The Role of Population Genetic Theory and Molecular Ecological Methods in Screening and Maintaining Healthy Freshwater Fisheries.

1.4.1 The 'Stock' Concept

Prior to the revolution in the wide scale applicability and application of molecular genetics and biochemical markers to screen for genetic variation, the concept of a fisheries stock was based upon the demography, geography and traditional ecology of a managed species or biota. Variables such as differential recruitment, mortality rates, geographical distribution and morphological variation were all utilized to delimit a stock (Carvalho and Hauser 1994). As an abstract definition, Ward (2000) quotes Ihssen et al.'s (1981) definition of a 'stock' as being representative of the general consensus within the fisheries science literature: "a stock is an intraspecific group of randomly mating individuals with temporal and spatial integrity." In other words, a stock comprises a deme, replete with the concomitant phenotypic and genetic properties they contain, the result of historical and contemporaneous evolutionary and demographic forces. It can be argued that this definition makes the case, albeit implicitly, for the inclusion of genetic information in determining what subset of individuals within a wider population constitutes a stock for individual management. Ihssen's biological concept of a stock de-emphasises but does not refute the importance of a 'sustainable harvest' (in terms of life-history characteristics) (e.g. Gauldie 1988) or the idea of a stock delimited by geographic location (e.g. the 'fishery stock' concept of Smith et al. 1990). Also, although it emphasises genetic integrity, Ihssen's concept does *not* assume strong genetic isolation, unlike alternative isolationist concepts (Carvalho and Hauser 1994).

A great deal of the early debate on the use of a genetic stock concept centred upon the problem of marine species, understandably so given the commercial over-exploitation of marine resources. Marine fish exhibit, in general, very high levels of within-population genetic variation and little, or undetectable, levels of structuring among populations compared to freshwater and anadromous species (DeWoody and Avise 2000). Thus, the genetic conventions of a biological stock concept based upon degree of reproductive isolation are less problematic for those freshwater species for which significant geographically-based spatial structuring is expected, at least among hydrologically distinct catchments. Thus, the use of the word 'stock' in this thesis shall follow Ihssen's definition, unless where stated otherwise.

1.4.2 Advances in Molecular Genetics Techniques for Screening Spatio-Temporal Genetic Diversity.

Although studies of the population genetic structuring of fish (marine, freshwater and diadromous) began in earnest with the rise of the allozyme analyses in the mid-to-late 1960s, the idea of utilizing genetic markers was born decades earlier (Utter 1991, 1994). In the 1930s an attempt was made to detect blood group differences in populations of the herring *Clupeidae* spp, although to no avail (De Ligny 1969). However, it was only when the use of allozymes[4] became widespread, that it was possible to identify, genetically, population subunits within species and the existence of stocks through specific stock assignment or SSA (Carvalho and Hauser 1994). Allozyme studies flourished, although doubts were raised as to their general applicability due to the possibility of natural selection operating at some allozyme loci (Powers et al. 1991). Concerns were also raised that allozymes may not reveal all the variation present in the protein-coding loci due to the possibility of synonymous (like-for-like) changes of amino acids (due to the degenerate nature of the genetic code, at the third, and less so at the first, codon position) that result in equally mobile proteins on a gel; that is to say, the low expectation of the number of alleles per locus limits the statistical power of some analyses (e.g. the exact-tests of genotypic differentiation among populations) (Estoup and Angers 1998)).

With the 1970s came the nucleic acid revolution. DNA extraction procedures and the discovery of restriction enzymes[5] ushered in a new era of screening for genetic variation at the level of the nucleotide sequence. Particularly appropriate for such screening was the mitochondrial genome (mtDNA), which could easily be extracted from cellular detritus; it also, as it turned out, contained many potential restriction sites. The presence or absence of restriction sites at particular loci within the circular genome, and the resultant banding patterns visualised on agarose gels, were used to configure haplotype relationships among individuals and among populations (Avise et al. 1979a). A haplotype is a length of DNA inherited as a single unit (mtDNA is inherited through the matriline as a 'single' copy, i.e. it is 'haploid'). The differing frequencies of haplotypes between drainages, for example, had obvious implications for ascribing management status, particularly at broad-scales,

[4] Mendelian-inherited enzyme variants, the alleles of which (the 'stained' protein variants themselves) can be discriminated from one another via differences in mobility through a starch gel due to differences in charge, mass and/or tertiary-configuration. The differences in gel mobility arise ultimately to amino acid substitutions, themselves the result of nucleotide substitutions within the coding regions of the loci from which the enzymes were transcribed and translated.

[5] Restriction endonucleases cut double-stranded DNA at specific palindromic nucleotide sequences of 4-6bp in length (Meselson and Yuan 1968).

where the comparatively high mutation rate of mtDNA (versus nuclear DNA) and its four-fold lower 'effective size'[6] proves to be particularly informative over long-time scales (over which different populations have diverged significantly so that haplotype frequencies in different populations have drifted apart to statistically significant levels). The resolution of nucleic acid screening approached its apogee with the advent of the direct sequencing of stretches of DNA (Sanger et al. 1977; Maxim and Gilbert 1977). Now every single base within a 'gene' was theoretically open to direct scrutiny for differences between individuals. This situation was greatly improved upon with the discovery of the polymerase chain reaction (PCR) by Muller (Saiki et al. 1985). PCR allows the exponential amplification of extremely small amounts of DNA. During the same period in the 80s, Jeffreys et al. (1985) discovered satellite DNA not involved in coding for protein or RNAs, tandemly repeated at loci found across entire genomes. Screening multiple loci, each varying in the number of repeated elements, provided a genetic 'fingerprint', which had many applications in forensics and organismal biology. These minisatellites, although useful, were dominant markers, and therefore had limited utility in determining substructuring of populations and concomitant estimates of population connectivity. However, the situation changed with the discovery of a related set of DNA repeat elements, the smaller, more ubiquitous – and most importantly – co-dominant microsatellites.

Microsatellite DNA sequences (alternatively known as short sequence repeats (SSRs)) consist of iterated sequences of 2-6bp motifs (e.g. GAGTGAGTGAGT…or (GAGT)*n*) typically some tens – hundreds of base pairs in length (Beckmann and Webber 1992). Microsatellites, although consistently distributed throughout the genome of almost every living organism and presumed selectively neutral, are in some instances associated with chromosomal structuring and other chromosomal and genic functions (Chistiakov et al. 2006). However, the vast majority of microsatellite loci have no known function and are considered truly neutral markers. In the Japanese pufferfish *Takifugu rupripes* microsatellites make up some 1.29% of the genome (Edwards et al. 1998), and in the genome of the three-spined stickleback *Gasterosteus aculeatus* the dinucleotide microsatellite repeat (CA)*n* occurs once every 14kb on average (Peichel et al. 2001). These facts highlight the ubiquity of their numbers and spatial distribution within the genome, thus providing many potential markers – a locus will differ in the number of

[6] Due to its mode of inheritance (matrilineal and haploid) a haplotype will 'drift' to fixation in a length of time that is four-fold less than the time it would take a nuclear allele to 'drift' to fixation from the same initial frequency as in the original (reference) population, assuming no selection and an infinite population size.

repeat units. They are also, generally, highly polymorphic due to rapid rates of mutation (between 10^{-2} and 10^{-5} per locus per generation)[7] (Ellegren 2000a) and thus have the potential to discriminate between populations due to differences in allele frequencies at multiple loci. Additionally, microsatellite alleles can be discriminated by size (distance travelled through an acrylamide gel) and heterozygotes differentiated from both homozygotes (the loci exhibit co-dominance). Additionally, unlike allozyme analysis, biological samples can be taken either from the animal, or from *ex vivo* samples such as shed hair, scales or fossil bone, without killing the subject. These properties have made microsatellites one of the most popular genetic markers in fisheries ecology and genetics, and throughout molecular ecology and population genetics generally (Chistiakov et al. 2006).

1.4.3 Identification of Population Subunits and Population Connectivity Using Genetic Data: Relevance for Stocks, Management and Conservation.

A stock concept built upon principles of 'spatial and temporal integrity' implies some degree of reproductive isolation and, by corollary, some degree of genetic divergence from each other. Consider a simple model of the fragmentation of a single ancestral population. For these daughter populations to become genetically divergent depends upon a number of factors: pre-separation demographic history (population coalescent time and parental population effective size (a descriptor of the effect of the random sampling of reproducing individuals in the ancestral population)), time since separation, demographic history since parental fragmentation (population expansion, stasis or contraction), rates of exchange of reproductive individuals among the daughter populations, presence and strength of local selective regimes and deviations from non-random mating. Each of these events may by themselves cause evolutionary change, but often occur together in various proportions and over a spectrum of intensities. The likelihood that two or more daughter populations should diverge by the exact same degree of differentiation is negligibly small, thus biodiversity is created. The primary aim of contemporary conservation and management genetics - fisheries included - is the preservation of genetic diversity, but for what reasons?

[7] Probably as a result of 'slippage' during DNA replication, a single-stranded DNA strand loops up on itself, resulting in one strand gaining a repeat(s) and one strand losing a repeat(s) when the complimentary strands are lengthened. Models of microsatellite mutation and evolution are important parameters in their own right, and shall be considered in the next chapter.

Ryman (1991) provides a comprehensive overview of the goals of the conservation genetics movement and its implications for management. One of the issues is purely from a humanistic standpoint: it's an ethical responsibility of mankind to preserve what he might destroy (Riggs 1990). Second, our knowledge of species, populations and their genes is limited, and preserving this diversity is our only way of achieving greater knowledge, which may be of benefit to mankind in the future. It is to this end that many authors strive to advocate the preservation of all diversity (Ryman 1991 and references therein). Rather more prosaically, however, perhaps the most compelling reason to acquire knowledge of - and preserve - diversity from a management and fisheries perspective is that it allows for further exploitation of a particular fishery.

The conservation of all biodiversity has been advanced in terms of maintaining the long-term potential for further evolutionary adaptation to changing environments. The degree to which genetic divergence among populations is adaptive is unknown, but morphological variation is often taken as a logical proxy (Allendorf and Luikart 2006). However, fish generally exhibit lower narrow sense heritabilities[8] than other vertebrates in their morphological traits (Purdom 1993) and as such individuals may be more adept at applying phenotypic plasicity to new environments and circumstances. However, this application of narrow sense heritabilities may overlook the significant heritabilities associated with phenotypic plasticity and metabolic pathways, and the importance of each cannot be understated. The upshot is that the translocation of fish from one part of its range to another is no guarantee of its continued existence and as such as much genetic variation should be conserved so that, by extension, a significant part of the species' heritability and therefore a continued ability to 'evolve', be preserved (Carvalho 1993).

Most genetic studies of fish, however, utilize selectively neutral markers. This neutral variation, often first assayed as levels of heterozygosity, is taken as a proxy for levels of genetic variation over the genome as a whole. Because most adaptive evolutionary change is continuous and polygenic, then the nearest proxy in the genome, without actually screening for loci that may be under selection, are multiple neutral marker loci (Allendorf and Luikart 2006) (although this situation is rapidly changing with the genome wide screening of EST-linked microsatellites or SNPs, single nucleotide polymorphisms). Such a viewpoint values high scores of heterozygosity in neutral markers as epitomizing the genome in general, a corollary of which should be a positive correlation with fitness

[8] Most adaptive evolutionary change occurs via the additive effects of individual loci. Narrow sense heritability is a metric that describes the proportion of the phenotypic variance due to such effects.

(Allendorf and Leary 1986; Hansson and Westerberg 2002). Furthermore, this variability found both within and among populations provides, in reserve, the potential for continued colonization of environments via adaptive potential (Carvalho 1993) and a source of re-colonizing migrants to these areas (Hodder and Bullock 1997).

Potamodromous fishes often inhabit rivers in which habitat patchiness is commonplace (Matthews 1998). If, as is the case in many freshwater fishes, populations within species take advantage of this heterogeneous environment, and spawning is limited to geographically separated patches, then one should expect some degree of reproductive isolation and genetic divergence. This divergence may reflect local adaptation to particular elements of the patchy habitat (Coelho and Zalewski 1995). In the face of the possibility of local extinction, depending upon the degree of isolation and adaptive divergence of resident populations, then one should apply caution to restocking from genetically different populations from other stretches of river, or from farmed fish whose genetic constitution may have been severely altered through captivity (Templeton 1986). Stocking from farms may also do nothing to combat the genetic erosion of a species if single farms are used and individuals are distributed about the species' range. In many cases, fish may be transplanted from different drainages where populations have diverged without the exchange of any migrants for tens of thousands of years. In such cases, whole co-adapted gene complexes may, in a single act of cross-fertilization, be destroyed if the introduced fish breed with the endogenous stock. In the event of this outbreeding, the F1 hybrid progeny may be less fit than either parent and the introgressive translocation would fail in the long term, especially if the introduced parents fare less well in the new environment than the old. Such instances of outbreeding depression (Templeton 1986) may be more common in freshwater fauna due to the physical and temporal separation between populations among drainages.

Population genetic information has gradually been incorporated into the fisheries management literature. As a result of this synthesis, operational definitions of what constitutes a 'stock', or 'stocks', to merit individual management were derived. The term 'management unit' has not coalesced into a single working definition, however most definitions involve some genetic criteria. Moritz (1994) denotes MUs, on genetic principles, as single, integral groups of individuals that possess significant divergent frequencies of alleles at multiple loci. As Palsbøl et al. (2006) note, the phrase 'significant divergence' has since come to mean deviation from panmixia. The deviation from panmixia model makes the classification of separate drainages in freshwater organisms as

MUs a relatively simple exercise if based upon a simple demographic definition of panmixia. However, such designations need to be backed up by other evidence (e.g. genetic divergence and/or ecological characteristics). Palsbøl et al. redefine MUs based upon not just the rejection of panmixia, but upon the degree of genetic divergence, asserting that the delineation of MUs is problematical if dispersal is low (but one cannot reject panmixia due to low statistical power), or dispersal is high between subpopulations with the attendant power to reject panmixia. There is not a one-size-fits-all model to assigning MUs over all taxa (Palsbøl et al. (2006). However, it has been suggested that a level of dispersal of less than 10% (above which populations become genetically mixed (Hastings 1993)) among subpopulations may be a reasonable indicator of a breakdown in panmixia. As a caution, the assignment of MUs should always be embedded in the biology of the species in question, especially taking into consideration other lines of evidence as to the extent and nature of dispersal. The derivation of dispersal metrics (rates of migration) and measures of genetic divergence, and the difficulties incumbent in their proper interpretation, shall be introduced in the next section on theory and methods

1.5 Study Aims and Objectives.

The literature concerning the distribution of genetic diversity within and among populations of coarse fishes is small. Much of this reticence to assess coarse fishes for spatial genetic information arises from skepticism regarding a long and complex interrelationship with *Homo sapiens*, stretching far back into pre-Historical times, at least in the UK. Many have assumed that anthropogenic movements will have muddied the waters as regards overall levels of genetic divergences, thus negating the effectiveness of performing surveys and assuming little gain to be had for the mining of genetic data for their application to management schemes. This view needs to be tested, less it be too widely applied to fish species that possess quite different natural and anthropogenic histories, even within the coarse fishery.

Thus, this thesis has two main, complementary aims:

Firstly, to apply current molecular genetics techniques and theory to elucidating the degree of divergence among roach populations from a number of English rivers that encompass a broad range of catchments in which the roach is part of the native piscifauna (Chapter Three). This broad scale analysis of genetic diversity is complemented by a second study,

adopting multiple microsatellite marker loci, analyzing the micro-structuring of roach populations within two contrasting rivers (which differ in their hydrology and anthropogenic alteration) in the southeast of England: the River Thames and the River Suffolk Stour (Chapter Four).

Second, the studies of diversity shall be presented not only in the context of local and nationwide fisheries management, but also in the context of the roach's natural history and ecology, information from which can be used to further refine and implement more effective management strategies and the delimitation of appropriate MUs.

Whilst coarse fish have been studied at the level of population substructuring before, as indeed has the roach, these studies have been few and far between and often limited in their scope. This study, for the first time, looks in detail at the causes of genetic differentiation within and among populations of roach at two hierarchical scales.

Chapter Two – On Theory and Method

"I am not a mathematician at all. My way of reading Sewall Wright's papers, which I still think is perfectly defensible, is to examine the biological assumptions the man is making and to read the conclusions he arrives at, and hope to goodness what comes in between is correct." **Theodosius Dobzhansky on Wright (Oral History Memoir, 1962), quoted from Provine (1989).**

2.1 Introduction

The purpose of this chapter is to outline the basic and fundamental theoretical and methodological foundations upon which this thesis is built. It will introduce the nature of genetic variation, models of evolutionary change (beginning with the evolutionary antithetical model of Hardy-Weinberg), the checks and balances of drift and migration/mutation in contributing to spatial, temporal and intra-population genetic variation, and statistical metrics (both frequentist and model-based) of genetic divergence and of migration. The chapter finishes by introducing both the microsatellite and the mitochondrial marker in population genetic studies. Laboratory methods and specific analyses, and their justification, will be outlined in detail in the relevant 'materials and methods' sections of their respective chapters.

2.2 Population Genetics Theory

2.2.1 The Raw Material of Evolution: Mutation and the Genesis of Variation.

Evolution simply cannot proceed unless there is variation present among individuals. Further, adaptive evolution cannot proceed unless there is variance among individuals in the part of the genome that contributes to a selectable phenotype in an additive fashion (Wade 2006). Over evolutionary timescales, mutation is the prime source of variation within and among populations (migration becomes an analogue of mutation over

ecological and temporally shallow timescales). Genetic mutations may be small in effect (e.g. the substitution, removal or addition of single bases to a segment of DNA) or gross (e.g. chromosomal rearrangements); neutral (e.g. nucleotide substitutions in the 3rd base position of a triplet codon result in a functionally and polymerically identical protein: so-called 'silent' or 'synonymous' mutations), non-neutral (e.g. a base substitution in the first or 2nd codon position that results in an amino acid change may affect the functionality of a protein)[9] or nearly-neutral (in which the selective value of the mutation is a function of the population size in which it finds itself (Ohta 1973, 1992)). This study purposefully uses markers in which neutrality is an important conceit, so as to uncover past processes and the demographic dynamics of present populations, without the complications arising from non-neutrality (e.g. selection or mutational drive). Therefore, the following brief exposition shall consider only those mutations that have no selective effect. Once a mutation appears in a finite population it will become fixed in, on average, $1/2N_e$ generations (see pages 51-52 for a discussion of this Sewall Wright effect or genetic drift). New mutation must occur to maintain variation at a locus that is approaching fixation. Mutation at a locus occurs with frequency $2N_e\mu$. Therefore the average allelic variation of a locus at any particular time is simply the product of the mutation rate and the probability of fixation of a mutant at that locus, $(1/2N_e)\bullet 2N_e\mu = \mu$ (Kimura 1968; King and Jukes 1969). This equation describes the amount of variation present at mutation-drift equilibrium.

Mutation rates may vary among loci and over time (Kelly and Rice 1996). The equilibrium of drift and mutation may be at various stages of imbalance or balance in different parts of the genome (which may have had an important effect on genome organisation over evolutionary timescales (Lynch and Conery 2003)). Potentially, every nucleotide base is a locus of interest, especially if this interest involves phylogenetic reconstruction. The probability of a mutation occurring within the purine or within the pyrimidine group of bases is often of the order of four-fold that of transversions (purine to pyrimidine or vice versa) especially in animal mitochondrial DNA in which transition:transvertion ratios are particularly skewed (Belle et al. 2005). Mutation rates may also be elevated in distinct hotspots that may be associated with errors in the DNA replication machinery (Rogozin and Pavlov 2003). Each base position has the potential to vary in the rate of substitution at that site. This rate heterogeneity is an important factor that needs to be addressed in

[9] Dependent upon the abiotic, biotic and epigenetic environment in which the mutation (via the phenotype) finds itself, it will either be invisible to selection (effectively 'neutral'), or have a negative or, less likely, a positive consequence for the carrier. The quantitative measure of this effect is called the selection coefficient, *s*, and by definition must affect the fitness of the individual (*w*) in which it resides.

phylogenetic reconstruction. Rate variation over a given sequence can be modelled by the gamma-distribution. The shape of this distribution (α) determines whether the effect of rate heterogeneity is severe with large effect or whether it is moderate or unimportant. This statistic is often used in conjunction with mathematical assumptions about rates and the direction of mutation between all four classes of nucleotide, so as to model the evolution of nucleotide substitution at particular loci of interest. The first and most basic model, that of Jukes-Cantor (1969), assumes no rate heterogeneity and that the probability of one base mutating to another is equal among all sites at all times. All subsequent models are an extension and extrapolation of this simple predecessor (see texts by Graur & Li 2000 and Nei & Kumar 2000 for a full exposition of the many interrelated models).

As just discussed, models of evolution are important in the interpretation of genetic data, especially so in model-based analyses. Summary statistics are also very useful, particularly in respect to measures of genetic divergence based on frequentist principles (e.g. F_{ST}, see p55). A basic measure of genetic diversity is the mean expected gene diversity (for nuclear loci) or the mean expected nucleotide diversity (for mtDNA) (Nei and Li 1979). These measures can be computed directly from frequency information and are not sensitive to the limits of population size (Graur & Li 2000). Both of these summary metrics summarise the probability that two randomly drawn alleles, or haplotypes, are different from one another. In the case of nuclear polymorphism, average allelic diversity (heterozygosity), *H*, is given by:

$$\frac{1}{L}\sum_{i=1}^{L} h_i$$

Where L = the number of loci and h_i is the gene diversity at locus *i*. For nucleotide diversity, the analogue, Π, is given by:

$$\sum_{ij} x_i j_i \pi_{ij}$$

Where x_i is the nucleotide frequency of the *i*th individual and x_j likewise for the *j*th individual. π_{ij} is the proportion of nucleotide dissimilarities between the two sequences.

As previously mentioned, finite population size has an impact upon the amount of genetic variation present in a population. This is chief among many factors that may affect the distribution of genetic variation within a population However, the baseline model of genetic variation within a population from one generation to the next invokes no evolutionary change at all: the Hardy-Weinberg model of evolutionary stasis.

2.2.2 The Baseline Model of Evolutionary and Population Genetics: The Hardy-Weinberg Equilibrium.

The null model of evolutionary genetics is the Hardy-Weinberg equilibrium model, independently discovered in 1908 by the English mathematician G. H. Hardy and the German physician Wilhelm Weinberg (but see Provine (1987) for earlier deductions of the same principle in the United States). The model, in its simplest form, states that in the absence of evolutionary forces (such as natural selection, mutation, migration and non-random systems of mating), and in populations of such size that the effects of statistical sampling are negligible, a population from one generation to the next shall experience no change in the frequencies of its alleles at a given locus: that is to say, evolution has not occurred (allele frequencies are said to be in Hardy-Weinberg equilibrium (HWE)). The model rests upon assumptions not usually commonplace in nature, but it provides a useful test, not only as an initial indicator of the influence of evolutionary phenomena, but as an indicator of the suitability of selectively neutral marker loci for studies of population demographic phenomena such as rates of inter-population migration and the elucidation of effective population sizes. HWE may be found wherever loci are inherited in a strictly Mendelian fashion and when the above criteria for evolutionary change are not met. Thus, whilst some loci with a direct phenotypic effect may influence the selection of mates, for instance, other unlinked loci 'blind' to the influence of natural or sexual selection, and given the relaxation of the other evolutionary forces, will closely follow Hardy-Weinberg expectations of allele frequencies in subsequent generations. HWE can be destroyed in a single generation whenever one or more of the above assumptions are violated; and, conversely, restored in a single generation of random-mating in a large population where drift, selection and migration have ceased to have an impact.

In molecular ecological studies, tests for conformity to HWE are routinely implemented when analysing suites of co-dominant molecular genetic markers. Whilst it may seem an unrealistic expectation that real world "populations" will exhibit HWE at all marker loci (the word "populations" is bookended by quotation marks to highlight the fact that

investigators often sample an area with little or no *a priori* appreciation of the actual demographic structure of an organism's range), the genomes of most eukaryotic species are large enough to yield literally thousands of potential markers (Koskinen et al. 2004). Aside from selection, and the sampling of individuals from pooled reproductive units, perhaps the biggest biological, and methodological, impediment to HWE is the census sizes of the populations being sampled. In finite populations within which low numbers of individuals are sampled, rare alleles will normally be found in heterozygous individuals, but the Hardy-Weinberg formula, which is based upon the binomial distribution of allele frequencies, will determine an unrealistic proportion of homozygotes (Allendorf and Luikart 2006), as the expected number of heterozygotes is increased by an amount proportional to a decrease in the homozygote categories (Levene 1949). As an example to illustrate the problem of sample size and the statistical methods to determine conformity to HWE in molecular ecological studies, the next section will focus upon a single microsatellite locus from the roach, Ca17.

The locus Ca17 is a tetranucleotide microsatellite locus with a basic repeat motif of $(TAGA)_n$. 40 individuals from the River Witham, in the East of England, were assayed, via PCR-based amplification, and visualised on an acryladmide gel. Two alleles were detected – one allele consisting of 228bp, the other possessing a length of 240bp. Thus, three genotypes are potentially extant in this population, and from an analytical point of view, 1 degree of freedom is available for traditional statistical analysis (χ^2-test). Table 2.1 describes the frequencies of the three genotypes, the inferred allele frequencies (according to the Hardy-Weinberg theorem) and the expected genotype frequencies given the actual frequencies of the two alleles. Finally, a χ^2-test is conducted to determine whether the scored genotype frequencies for the Witham population are within the expected range as predicted by the allele frequencies in a scenario of HWE. The χ^2 statistic is given in the bottom right hand panel of Table 2.1. This figure – 0.11 (highlighted in yellow and underlined) – indicates that there is no significant deviation from expected Hardy-Weinberg frequencies of each recorded genotype of CA17 in the Witham; that is the locus conforms to Hardy-Weinberg expectations. Generally, the level of significance (α – level), below which any deviation from expected HWE is greater than would be expected by chance alone, is given as 0.05, although this figure is arbitrary. However, note the expected number of the rare homozygote - it is less than one. The suitability of the simple χ^2-test is compromised when the expected numbers of any of the genotypic classes becomes too low, with limits in the literature set at varying numbers of individuals (five (Allendorf and Luikart 2006; Hedrick 2005a); three (Cochran 1954); and even as low as

one (Lewontin & Felsenstein 1965)). However, Hedrick (2005a) suggests that adjacent classes be combined when such individual categories number less than 5 (although this is not possible for a locus with just two alleles as the degrees of freedom will be zero).

Table 2.1: The determination of Hardy-Weinberg equilibrium for the locus Ca17 in a sample of 40 roach from the River Witham.

Genotype	Observed (O)	Expected (E)	$(O-E)^2/E$
228/228	30	29.75	0.002
228/240	9	9.49	0.025
240/240	1	0.75	0.083
Σ	40	39.99	0.11

Additionally, the χ^2-test loses power to detect subtle deviations from HWE, due to its conservative nature, in small sample sizes (< 50 individuals) (Hedrick 2005a). Many of these concerns regarding the application of χ^2-tests to determine conformity to HWE are particularly pertinent to studies of population genetics, due to the limitation of sampling rare or elusive organisms, the patchiness of organismal distributions and the unknown nature of underlying population substructure. Also, the marker loci used in such studies tend to exhibit greater levels of polymorphism than the example of Ca17 above. Many microsatellites loci are highly polymorphic, although most alleles will be rare in comparison to a few common ones. Thus, with loci possessing multiple alleles the effects of small sample size are particularly acute.

Exact tests (Raymond & Rousset 1995a; Weir 1996) are used in the majority of studies in which sample sizes are low, or vary among locations, and where the number of alleles, and hence, genotypes, are high. They are implemented in a range of population genetic software, such as GENEPOP (Raymond & Rousset 1995b). Exact tests are a useful alternative to the χ^2-test, not only for the increased statistical power, but are also more tractable when dealing with myriad genotypes. Exact tests, or exact probability tests, determine the probability of each of the possible genotypes given the frequencies of each of the alleles in a given sample (Hedrick 2005a). As allele numbers increase, simulation is often necessary to compute the likelihood that genotype frequencies differ from expectation by chance alone (Hedrick 2005a). The programme GENEPOP (version 4.0) (Rousset 2008) utilises the Markov Chain algorithm (Rousset & Raymond 1997), in which a series of random variables, genotypes in this case, whose values are predicated on the previous simulation in the chain (Beaumont & Rannala 2004), are used to maximise the probability of such an array of genotypes occurring by chance. One can also determine the fixation indices of alleles within a population (F_{is}), which are informative because they

indicate whether a population has a deficiency of heterozygotes (positive values of F_{is}) at a particular locus, or whether there exists an excess (negative values of F_{is}). The proportion of heterozygotes to homozygotes may be biologically informative, depending upon the patterning of heterozygote deficiency among loci and populations. As an extension of the previous example, consider Table 2.2 that shows the results of exact-tests (as implemented in GENEPOP) for the locus Ca17 in nine populations.

Table 2.2 indicates that each of the 9 populations is in HWE for the Ca17 locus (the Yare and Kent Stour are insignificant after Bonferroni correction). The *p*-values indicate that the null hypothesis (of HWE) should not be rejected for each of the populations and across all populations (p = 0.2314). All but two of the populations show an excess of heterozygotes, as predicted from theory (Levene 1949), for small populations (mean population size = approximately 33 individuals).

Table 2.2: Exact tests for conformity to HWE of the Ca17 locus in 9 populations

Population	Census size	*p*-value*	Significant?	F_{IS}
Great Ouse	56	1.000	No	-0.091
Witham	40	0.548	No	0.064
Yare	30	0.026	No	-0.073
Yorks Ouse	8	1.000	No	-0.077
Kent Stour	44	0.015	No	0.065
Sussex Ouse	22	0.143	No	-0.355
Tees	36	1.000	No	-0.103
Trent	29	1.000	No	-0.061
Rhône	35	0.568	No	-0.214

* Bonferroni adjusted α = 0.0056; ‡ Weir & Cockerham 1984.

Overall, this locus displays only five alleles in 300 individual fish, with most populations diallelic for the two alleles mentioned above (228 and 240). The exceptions are the Yare (possesses private alleles 232 and 248) and the Kent Stour (private allele 236).

Table 2.3 outlines the various factors, both biological and systemic, that can have an effect on levels of heterozygosity in a population associated with deviation from HWE. Some of the factors have already been discussed (e.g. the Wahlund effect – the apparent loss of heterozygosity due to unbeknownst pooled sampling of two or more sympatric, yet

reproductively isolated, populations). Many of the factors are potential systemic errors associated with analysing molecular markers (mis-scoring of alleles) or not taking into account the vagaries of mutation (e.g. the presence of null alleles) or the visualisation process (allelic drop-out). Null alleles can have a major negative impact on studies of parentage analysis (Dakin & Avise 2004), population assignment tests (Carlsson 2008) and population structure (Chapuis & Estoup 2007) (but see Chapter Four).

Table 2.3: Potential causative factors for the deviation of populations from HWE as determined by their effects upon heterozygosity (taken from Hedrick (2005a)).

Effect on heterozygosity	Cause
Decrease	Selection against heterozygotes Inbreeding Assortative mating Gene flow of zygotes Wahlund effect over space or time Null allele(s) Allelic drop-out
Increase	Balancing selection of heterozygotes Outbreeding Disassortative mating Gene flow of gametes PCR artefacts of new alleles Mis-scoring of alleles

To complicate matters further, a combination of the above listed influences may interact in such a way as to cancel each other out, resulting in genotype frequencies consistent with those expected from HWE (Workman 1969). For example, the sampling of two reproductively isolated populations in communal feeding grounds will decrease heterozygosity over the entire population sample, as will genetic drift from small sample sizes, or limited sampling, whilst, conversely, gene flow from immigrants will increase levels of heterozygosity (Wang 2005).

2.2.3 Statistical Correction for Multiple Tests

It is worth noting at this juncture the problem of comparing multiple independent and pair-wise tests and the impact this has on overall levels of significance. Following Moran (2003), in a table of statistical comparisons there is a probability:

$$1 - (1 - \alpha)\, N \qquad (1)$$

of finding one or more associations that are significant by chance alone (where α is the type I error level and N is the number of tests). Taking the contents of Table 2.2 as an example, there are 9 tests being made with regard to HWE, and thus a probability of approximately 0.45 of achieving a statistically significant deviation by chance. To counter the appearance of fluke results, one administers the Bonferroni correction (Rice 1989), such that the α-value is adjusted by dividing it by a value *n*, where *n* is the number of tests (in this case 9 – they are not pairwise). Thus the adjusted α-value for 9 independent tests is 0.05/9 = 0.006. However, where we have direct comparisons of particular parameters (e.g. a measure of population divergence such as F_{ST})) between a number of categories (e.g. the 9 populations of Table 2.2), then the number by which the α-value should be divided is given by:

$$k\,(k-1)/2 \;=\; n \qquad (2)$$

Or alternatively, the revised α value is given by:

$$1 - (1 - \alpha)1/k \qquad (3)$$

where k represents the number of categories (e.g. populations). For k = 9 the equation (2) resolves as 36, which equals the number of possible pair-wise comparisons of 9 sampled populations. The α-value (e.g. 0.05) is then divided by 36, in this example, to attain a figure of 0.00138, also given by equation 3 (Sokal and Rohlf 1995). Therefore, to achieve significance at α = 0.05, actual *p*-values of population divergence need to be ≤ 0.00138. However, the use of the simple Bonferroni correction has been criticised as being too conservative by some authors, lacking in statistical power to detect subtle deviations from null distributions (Moran 2003; Nakagawa 2004; and Narum 2006). Some authors have criticised – and developed alternatives – to the standard Bonferroni correction (e.g. Holm

(1979)), however the standard correction is used in this study as the quality and quantity of included data should allow for a robust detection of non-random deviations from chance.

2.2.4 Linkage Disequilibrium

Each marker locus used is an independent source of data providing that they all follow Mendel's second law of inheritance, that of independent assortment. However, this 'law' may be broken in two respects, one that is fatal to the notion of independence and the second that is potentially useful if the contravention is not universal among associated loci in all populations. Close physical linkage on the same chromosome of two or more loci will cause a non-random association of alleles among the loci between generations due to their inheritance in close association (given a low rate of chromosomal recombination between them)[10]. If this is the case, this linkage disequilibrium (Lewontin and Kojima 1960) should be apparent in all populations. Thus, linked loci must be whittled down to a single marker – leaving for analysis the more polymorphic one. The second case of disequilibrium results in a similar non-random association of alleles, although this time the cause is not physical linkage on the same chromosome, but a statistical association of alleles from populations with near-historical or recent allelic differences that are sampled sympatrically. This phenomenon may be exemplified in the roach – a species that spawns after considerable migration to a number of spawning grounds, about which there is some evidence for fidelity of spawning individuals (Goldspink 1977; L'Abée-Lund and Vøllestad 1985), which then return to productive post-spawning, recuperative feeding areas shared with other 'spawning shoals'. The relative reproductive isolation of spawning shoals will result in their genetic divergence, and the sampling of populations in communal feeding grounds will result in the Wahlund effect (see above) and non-random associations of alleles within gametes (so-called gametic phase disequilibrium (Crow & Kimura 1970)). Normally, because non-physically linked alleles possess the maximum rate of recombination (e.g. independent assortment) within the germplasm, gametic disequilibrium (D) will break down by 0.5 every generation until the equilibrium frequencies of alleles has been reached. This would be the case if there were an initial one-off case of population admixture followed by random breeding. However, D can be maintained through a number of processes such as selection, population bottlenecks and by the maintained fidelity of spawning and shared summer feeding grounds. The existence of gametic disequilibrium in tandem with the Wahlund effect may indicate the summer

[10] Note: loci may be located on the same chromosome and not contravene the principle of independent assortment if the rate of recombination between them is maximal (i.e. $r = 0.5$).

feeding grounds of two or more reproductive spawning shoals, or demes. Disequilibrium can also be used to estimate effective population sizes (Waples 2006a), although such estimates may be confounded by the presence of sympatric demes in sampled areas. Linkage disequilibrium among loci within gametes may also result from the action of genetic drift in finite populations (see below).

2.2.5 Departure from Equilibrium Conditions

The most obvious deviation from the assumptions of the Hardy-Weinberg theorem is that of infinite population size. This deviation alone, as we shall see, is enough to facilitate evolution if the process is given enough time and there is the availability of allelic variation to begin with. Non-random systems of mating, the occurrence of mutation, migration between fragmented populations and natural selection will all cause changes in allele frequencies over the course of generations. If each of these forces acted alone (apart from constant mutational pressure), many alleles would approach fixation and evolution would grind to a halt. These forces do not occur in isolation from each other and they combine to cause fluxes in genotypic and phenotypic evolution, the scope of which depends upon their relative contributions. In certain circumstances equilibrium is reached (e.g. between the opposing forces of genetic drift, selection and mutation), the stability of which is governed by the many underlying properties of the genetic system under study and its prior evolutionary history. An example of a non-equilibrium condition concerns the mitochondrial genome (or loci therein) in small populations, which, because of its mode of inheritance, is liable to fixation in small populations, despite the actions of selection or migration. The signals of mtDNA haplotypes scattered around northern Europe in many eukaryotic taxa show a great deal of geographical patterning in terms of haplotype diversity because of historical bottlenecks associated with the glaciations of the last Ice Age. Populations became fragmented, and the isolates became fixed for different haplotypes such that these haplotypes and their descendants became ‘frozen’ due to the peculiarities of demographic expansion after the retreat of the ice sheets (see Chapter Three). At smaller spatial and temporal scales, the equilibrium processes of mutation, selection, genetic drift and migration predominate, although mutation is sometimes ignored due to the timescales involved. When one of these processes dominates over the others beyond the boundaries of the equilibrium, the equilibrium is broken and a population may drastically alter in its allele frequencies –resuming evolution.

2.2.6 Genetic Drift and the Effective Population Size.

Every population is finite. Thus, only one evolutionary force is universally pervasive: genetic drift. It interacts with all other known evolutionary factors, such that it determines the amount of genetic variation within and among populations (Wang 2005). Genetic drift occurs because of the error associated with sampling gametes from within a finite population. An extreme example for illustrative purposes is to consider a typical Mendelian monohybrid cross between two autosomal heterozygotes (e.g. Aa x Aa). Assuming sampling with replacement, i.e. the two parents have two offspring, there's a one in sixteen chance that either one of the alleles will be lost (Wayne and Miyamoto 2006). The population would then be homozygous for the other allele. The successful allele is present due to nothing other than caprice associated with Mendel's second law (i.e. the stochasticity inherent from the binomial sampling of a limited number of gametes). Here we have a definition of drift relating to the loss of heterozygosity within a population, or, alternatively, the accumulation of homozygosity. Assuming that all other evolutionary factors are absent, heterozygosity will be lost from a population at a rate that is inversely proportional to population size, so at time t+1 heterozygosity (*H*) will be:

$$\left(1 - \left(\frac{1}{2N}\right)\right) Ht$$

Where *H*t is the heterozygosity of the population in the preceding generation (Hedrick 2005b). Therefore, genetic drift is most effective when population sizes are small. Genetic drift was first considered by R. A. Fisher (1930), but elaborated upon much more extensively by Sewall Wright (1931) who extended the idea into a number of productive avenues. Randomly mating populations of finite size, subject to drift, are termed Fisher-Wright populations in the literature. They represent a simple representation of a deviation from the Hardy-Weinberg model. The implementation of these models enabled theoretical work into the fate of genetic variation to be made more explicit. The major conclusion from the application of the Wright effect is that genetic drift tends towards weeding out genetic variation, particularly in small populations and those populations founded by a few individuals (the so-called founder effect (Mayr 1954)).

However, even large populations may drift towards fixation of alleles when these alleles are on no other evolutionary trajectory. Furthermore, historical demographic bottlenecks will purge populations of genetic variation, and leave a signature of reduced allelic

variation embedded within the genomes of such organisms. A severe decline in population size will have the general effect of removing the rare alleles within a population, thereby reducing allelic diversity. However, the amount of heterozygosity is reduced at a rate less than the loss of the number of alleles. Thus, ratios of expected:observed heterozygosity and allelic diversity may be diagnostic of prior demographic history, perhaps indicating populations at risk of genetic erosion and the effects of inbreeding, which are significantly increased within small populations (e.g. Bijlsma et al. 2001).

Genetic drift determines the amount of variation that remains within the population that may later be utilized by selection for adaptive evolution in the wake of demographic expansion or environmental change. Depending upon the genetic parameters under consideration, and the demographic properties of a population, genetic drift affects the properties of genetic variation within a population in different ways. The impact of drift may be measured by the increase in the average probability that homologous alleles will be reunited in an individual with reference to an initial grandparental generation (identical-by-descent or IBD)[11]. Thus, for homologous alleles, genetic drift increases the likelihood of average pedigree inbreeding in a population. The 'effective size' of the surveyed population is then the number of breeding individuals in an idealised Wright-Fisher population in which the same outcome of inbreeding would be attained (Wright 1931). This is the inbreeding effective size, N_{eb}. It is generally less than the census population size, but not always. Another variant of the effective size concept considers the variance in allele frequencies among generations, and is termed the variance effective size, N_{ev} (the definition of which is the same as N_{eb}, but the genetic parameter under consideration is not the increase in inbreeding, but the variance in allele frequencies (Crow and Kimura 1970)). These are the two main concepts of 'the effective population size' (Sjödin et al. 2005), although there are others (e.g. the eigenvalue effective size (Ewens 1982), relating to the loss of heterozygosity through the generations, and the coalescent effective size (Nordborg and Krone (2002)). These size concepts measure biologically different phenomena (Templeton 2006), estimates of which may differ in their values for a single population, depending upon recent demographic and evolutionary history. All estimates are based upon measures of variance: the greater the variance, the lower the effective size (Nunney 1995). N_{ev} and N_{eb}, the most commonly used measures, approach similar values when populations are at evolutionary equilibrium (Wang 2005), and unless where

[11] Note that alleles may be in the same homozygous 'state', e.g. 'AA' or 'aa' (identical-by-state, IBS) and still not be considered IBD. The two alleles are only IBD (autozygous) – and the individual truly 'inbred' for this locus - if both alleles were inherited, via related antecedants, from an individual in which the alleles coalesce. (Templeton 2006).

specifically stated, they shall not be distinguished. In simple terms: the effective size indicates the number of idealised individuals that contribute to the allelic diversity of the subsequent generation, given the genetic parameters of the real population undergoing genetic drift (i.e. not all individuals in a population contribute alleles to the next generation, and these 'non-breeding' individuals are discarded from contention).

In management and conservation circles, the effective size represents an important parameter, as it indicates the amount of genetic variation within a population, and genetic management and breeding schemes can utilize this information to maximise demographic growth without subsequent loss of genetic variability. Unequal gene flow among recently fragmented populations may lead to stochasticity in gene transmittance within the sub-populations, such that effective sizes are dramatically reduced (Palstra and Ruzzante 2008). Low effective sizes also indicate populations at risk of extinction due to associated inbreeding effects and possible loss of allelic diversity, which reduces evolutionary potential to adapt to changing conditions (Lynch et al. 1995; Frankham et al. 2003). However, measuring N_e is extremely difficult, unless explicit demographic and pedigree information is available for a number of generations for the same population. Also, N_e is based on the assumptions of the Fisher-Wright model, concessions to which are contingent on mathematical tractability (Sjödin et al. 2005). These simplified conditions are often not satisfied in natural populations**.**

The inbreeding effective size is often estimated by determining allele frequency flux over a number of reproductively independent generations (Jorde and Ryman 1995), although populations with overlapping generations have been given a theoretical footing (Nei and Tajima 1984; Pollak 1983), in which case one may measure the inbreeding effective size of 'breeders', i.e. of single cohorts (Beebee 2009). The so-called temporal method uses the harmonic mean of the change in allele frequencies over a number of given generations (Pollak 1983). Thus, a reduction in population size in just one generation (bottlenecking) will lead to a reduced effective population size; the number of contributing individuals is, in one generation, severely reduced (Luikart et al. 1999), impacting future generations until new variation is introduced from mutation and/or migration. These estimates of effective size are dependent upon temporal data being available and that it is the *same* population that is being sampled in each time period (there are also issues dealing with age-structured populations and that discrete age classes differ in their allele frequencies which need to be taken into account (Jorde and Ryman 1995)). One-shot measures of N_e

have been developed, although until recently (e.g. Beebee 2009) they have not been tested empirically.

2.2.7 Population Structure

Determining genetic discontinuities is a fundamental aspect of all management and conservation genetic strategies. That population structuring is an important component of the evolution of species was formulated by the work of Sewall Wright (1931, 1938a, 1938b and 1951) and followers. As a result (see explication in Chapter One) it has become desirable to conserve much of the variance in the distribution of genetic variation. In the molecular ecology literature there are two basic, frequentist methods to determine differences: exact-tests and the application of inbreeding coefficients. Exact-tests for allelic differentiation use the basic exact test formulation of Fisher (Raymond & Rousset 1995a), whereas genotypic differences are calculated by using an unbiased estimator of the log-likelihood (G) scores of an exact test (Goudet et al. 1996).

The *F*-statistics (inbreeding coefficients) are a series of hierarchical measures in which the variance in allele frequencies within a metapopulation are apportioned according to a structural hierarchy. Devised by Wright to describe the distribution of genetic variation in both domestic and wild populations (Wright 1943, 1946 and 1951), the *F*-statistics (F_{ST} in particular) have found utility across the evolutionary continuum. The inbreeding coefficient *f* is described by Wright as: "the correlation between homologous genes of uniting gametes under a given mating pattern[12], relative to the total array of these in random derivatives of the foundation stock" (Wright 1965). Malécot re-stated inbreeding coefficients as the probability that two autosomal alleles in a diploid zygote are identical by descent, that is to say that they were independently inherited from a single ancestral individual (IBD) (Malécot 1948). The *F*-statistics as we know them now were derived from the inbreeding coefficient to determine the correlation of reuniting alleles at a locus within individuals relative to the subpopulation (F_{IS}), within a sub-population relative to the total population (F_{ST}) and within individuals relative to the total population (F_{IT}). Each of these three metrics are interrelated and thus non-independent (Wright 1965). F_{ST} is the important metric when determining population subdivision, as it deals with the distribution of allelic variance that is due to differences between subpopulations relative to that of the total population. Its relation to the other metrics is:

[12] For neutral markers the system of mating assumed is random with respect to mate choice.

$$F_{ST} = (F_{IT-} F_{IS})(1 - F_{IS})$$

F_{ST} can also be posited in terms of genetic diversity indices (heterozygosity or 'gene diversity') within populations and the difference among populations:

$$F_{ST} = 1 - \frac{H_S}{H_T}$$

Where H_S is the amount of observed heterozygosity averaged over all subpopulations and H_T is the amount of observed heterozygosity expected over the entire population if the population was in HWE (Nei 1973). F_{ST}, as originally conceived by Wright, dealt with a strictly diallelic genetic model: if two populations are homozygous for the two different alleles then $F_{ST} = 1$; conversely, if the two populations were fixed for the same alleles then $F_{ST} = 0$. However, modern estimators handle multilocus data.

In terms of variance of allele frequencies, Weir & Cockerham (1984) derived this estimate of population structure:

$$\theta = \frac{Var(p)}{\bar{p}(1 - \bar{p})}$$

Where Var (p) is the variance in allele frequency and $\bar{p}$ is the mean allele frequency (here F_{ST} is estimated by θ). These are an example of many different estimators of F_{ST} that have been developed over the years (Holsinger and Weir 2009). Weir & Cockerham's (1984) F_{ST} estimator (θ) samples from the allele frequency distribution (genetic or evolutionary sampling distribution) derived from all surveyed populations, taking into account the sampling variance inherent in sampling from a limited number of subpopulations from across a species' range (itself composed of an infinite number of subpopulations founded from a single ancestral population) (Holsinger and Weir 2009). Nei (1973), however, devised the F_{ST} estimator G_{ST} (the coefficient of gene differentiation) that reflects the degree of genetic differentiation among populations by deriving the metric from the gene diversities of those populations being compared (which are themselves a subsample of the 'genetic sampling distribution') (Holsinger and Weir 2009). When mutation process become an important determinant in the spatial distribution of genetic variation, it is sometimes more pertinent to use an F_{ST} estimator that takes this rate into account. R_{ST} (Slatkin 1995) is an F_{ST} analogue that incorporates a particular model of the evolution of microsatellite length polymorphism. Divergence estimates are derived from the average sum of squares of allele size differences between compared populations (Michalakis and

Excoffier 1996). Other measures exist that relate to the differentiation of mitochondrial haplotypes (Φ_{ST}) (see Chapter Three) and the differentiation of populations based upon additive genetic variation in continuous phenotypes (Q_{ST}), amongst others.

However, all these estimates of differentiation have come under recent critical scrutiny. Jost (2008) rejects the mathematical assumptions of estimators that are reliant upon the distribution of genic diversities (apportioned to within and among population components, e.g. G_{ST}) and any metric upon which there is a dependence upon within-group heterozygosity (including most estimators of F_{ST}). Jost argues that these indices mistakenly assume that heterozygosity possesses an additive property when apportioned, independently, into the two subcomponents of within and between subpopulations (Jost 2008). G_{ST} will always be – incorrectly - zero if within population diversity is very high and population independence is absolute. G_{ST} may even decrease as population divergence increases. Jost's paper follows on from historical disquiet of diversity-based estimators (e.g. Nagylaki 1998; Charlesworth 1998; and Hedrick 2005b) stipulating that G_{ST} only works well with low gene diversity. As an alternative, Jost offers D (for differentiation) to the canon of estimators of genetic differentiation deriving measures for H_S and H_T. D assumes the relationship between H_S & H_T to be multiplicative rather than additive and that the partitioning of gene diversity of a population includes a new component: that of the decomposition of diversity amongst subpopulations (H_{ST}). Differentiation is determined by:

$$D = \frac{\left(\frac{J_T}{J_S} - 1\right)}{[(1/n) - 1]}$$

D is Jost's differentiation estimator, J_T is the gene identity of all the pooled subpopulations, J_S is the gene identity of a subpopulation and n is the number of subpopulations. This equation has the potency to rank populations based on the actual levels of genetic divergence and ranks divergence faithfully from 0 (undifferentiated) to 1 (fully differentiated). Applied within the context of Wright's original island model of population subdivision:

$$D = 1 - \frac{G_D}{G_S}$$

Where G_S is the probability of uniting two alleles from the same subpopulation and G_D is the probability of uniting two alleles from different populations (Jost 2008). However, the utility of old and new estimates engages continuing debate (Heller and Siegismund 2009**;** Ryman and Leimar 2009; and Jost 2009).

2.2.8 Models of Population Connectivity and Estimates of Gene Flow

Wright's (1940) original conception of the infinite island model (see Fig 2.1) subdivides a population into an infinite number of subpopulations of equal size each exchanging with one another, with identical probability, a proportion *m* of effective migrants (that is migrants who contribute genetic material to the recipient population in the next generation). The finite population sizes of the subpopulations implies the loss of heterozygosity, at a rate proportional to the inverse of $2N_e$, for each subpopulation with the result that individual loci would eventually become fixed if not for the renewal of variation through migration (leaving aside mutation). F_{ST} can be related to the loss of heterozygosity through genetic drift within each subpopulation thereby increasing the amount of variation among the subpopulations. Gene flow among the populations will counteract the effects of drift and so reduce differentiation.

Thus at migration-drift equilibrium:

$$F_{ST} = \frac{1}{(4Nm + 1)}$$

Therefore:

$$4Nm = \frac{1}{F_{ST} - 1}$$

However, the use of *F*-statistics to estimate Nm, (actually N_em - the population-scaled effective number of migrants) although once widespread, began to attract criticism both theoretically (e.g. Whitlock and McCauley 1999) and methodologically (Bossart and Powell 1998). Whitlock and McCauley's criticism focuses upon the unrealistic assumptions behind the island model (e.g. failure to account for mutation, selection and unequally sized populations (Gaggiotti and Excoffier 2000)). Also, migration among the subpopulations is assumed to be symmetric (Tufto et al. 1996; Bahlo and Griffiths 2000).

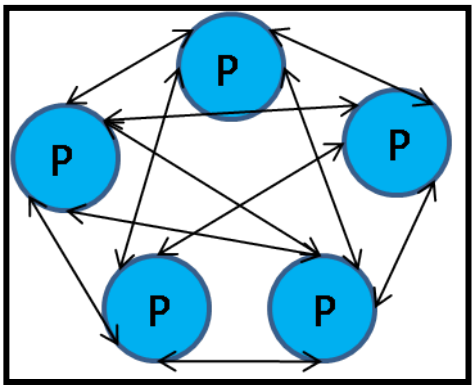


Figure 2.1: The infinite n-island model of Wright (1940). Each arrow represents the exchange of migrants, m, between each population (P). Each population is connected with every other population and exchange the same proportion of migrants with an identical probability.

The island model constitutes a 'metapopulation', but not in the classical sense as defined by Levins (1969); that of a set of subpopulations that inhabit heterogeneous habitat patches and each of which has the potential for individual extinction and recolonisation (see Hanski (1999) and Hanski and Ovaskainen (2000) for the impact of habitat heterogeneity on recolonisation and extinction). Thus extinction, founding events and the subsequent gene flow that occurs within the finite lifespan of individual populations all contribute to a heterogeneous genetic landscape (Whitlock 1992; Ingvarsson et al. 1997).

The island model does not possess an explicit spatial component (all populations are assumed equally distant from one another). In reality this assumption is almost always violated (unless there are just two populations). Simply put, the exchange of migrants, and by extension the degree of genetic divergence, now becomes a function of the distance between respective populations, a phenomenon known as isolation by distance (IBD) (Wright 1943, 1946). The influence of geographic distance is exacerbated when migration is limited to 2 or 1 dimensions: rivers and streams illustrating prime examples of a one-dimensional system for those organisms dependent upon the fluvial system for migration and reproduction. An early advance upon the island model was the one-dimensional stepping stone model of Kimura and Weiss (1964) (See Fig 2.2).

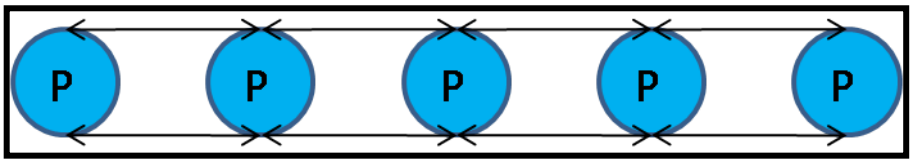


Figure 2.2: The stepping stone model of population genetic structure (Kimura and Weiss 1964). Each arrow represents the exchange of migrants, m, between each population (P). Although the probability of exchanging migrants with *contiguous* populations is identical, as in Wright's (1940) model, genetic differentiation at migration-drift equilibrium will occur due to the enforced spatial dimension to gene flow.

In this model, adjacent demes are most likely to exchange migrants, with the probability of exchange decreasing as a function of the distance between them. This model is seemingly a better fit for ecosystems such as rivers and streams. However, the directional flow of water currents will impact upon migration symmetry in many organisms (e.g. downstream gene flow determining greater genetic variability in those populations near the mouths of rivers (Hernandez-Martich and Smith 1997**;** Hänfling and Weetman 2006)). Sex-biased dispersal also presents a significant interpretative problem (e.g**.** Fraser et al. 2004), as does the presence of habitat patchiness (Pulliam 1988).

Other issues relating to the use of F_{ST} to determine Nm include incorrectly inferring on-going migration between populations when those populations have, in reality, been demographically isolated for generations. A large, highly variable population may be sundered into two daughter populations, but if genic diversity is high and drift has had too little time to effect significant differences in allele frequencies between them, then F_{ST} will erroneously infer continuing and homogenising gene flow between them. Similar expectations would be expected at the extremes of a species' range (Whitlock and McCauley 1999) where adaptation to novel, peripheral environments may be retarded by high levels of gene flow from within the range. Finally, estimates of N_em, calculated from F_{ST}, may underestimate actual rates of migration by 10% in highly fecund species in which the ratio of effective:census size is 1:10 (Frankham 1995). The calculation of Nm from F_{ST} also suffers from problems of variance (low F_{ST} scores give scores of Nm with greater variance): the relationship between F_{ST} and Nm is non-linear (Waples 1998).

The late 1990s saw the inception of new methods to determine indirect measures of effective migration. These methods reassessed the problem of migrant detection by employing model-based analyses: Bayesian analysis (BA) and maximum likelihood analysis (ML). Model-based methods utilise the concept of the coalescent, a theory developed to map genetic variation back through time (Kingman 1982a, 1982b). The

coalescent is the time-reverse of Malécot's idea of identity by descent (IBD): a coalescent event occurs when one traces the co-ancestry of two present day alleles back to the historical event which saw them descend from a single meiotic division. The coalescent examines the genealogy of a sample rather than modelling population wide fluxes in allele frequencies (Wilson and Rannala 2003). These methods do not make the assumptions of F_{ST} derived measures: that population sizes are equal and infinite in number, and, crucially, that migration is symmetric; but some do assume migration-drift equilibrium has been attained and for at least $4N_e$ generations (Beerli and Felsenstein 2001). As such, estimates based upon the coalescent are best viewed as long-term measures of gene flow (Weetman and Hänfling 2006).

A second model-based method is non-equilibrium in nature and provides an estimate of migration rate on the scale of three or so generations (Wilson and Rannala 2003). Such analyses are based upon Bayesian inference predicated upon the prior distribution of model parameters from which a posterior probability distribution is obtained to determine the likelihood of the model parameters given the data (multilocus genotypes). The method of Wilson and Rannala makes concessions for populations that are not in HWE. Both methods may be utilised in a complimentary way to investigate the exchange of effective migrants at different temporal scales.

2.3 Genetic Markers in Molecular Ecological Studies

2.3.1 Microsatellite Markers.

Microsatellite DNA markers, since their published debut in molecular studies at the beginning of the 1990s (e.g. Schlötterer et al. 1991), have become a genetic marker very much favored by molecular ecologists. In part this is due to the relative ease by which novel markers can be identified within the genomes of target species, but more importantly it is because of their mode of evolution: stepwise shifts in the number of repeat elements evolving alleles which are distinguished from one another by different numbers of repeat elements. Microsatellites tend to have rates of mutation that are several orders of magnitude greater than even the nucleotide substitutions of the mitochondrial control region (Ellegren 2004). This makes microsatellites useful for studying genetic and demographic processes on ecological timescales (Schlötterer 2000). Microsatellites are

high-resolution markers, apt for studies concerning rates and directionality of migration, microgeographic substructuring, etc. (Pena and Chakraborty 1994**;** Schlötterer 2004; and Selkoe and Toonen 2006). Microsatellite loci are scattered throughout the genome of most taxa (although, curiously, they are less frequent in birds than other vertebrates (Primmer et al. 1997)). Regardless of functional status, diversity at microsatellite loci is taken as being representative of the genome in general, and therefore indicative of the adaptive potentiality of genomic diversity (Väli et al. 2008). As such, measures of genetic diversity based on nuclear markers are used to assess adaptive longevity of extant populations. Because one can assess many allelic variants of microsatellite loci, they are especially sensitive to demographic and historical events that reduce allelic diversity. They are useful in addressing the concerns of inbreeding depression and the associated loss of fitness exacerbated by small census sizes. Their high variability also make them exceptionally useful in determining rates and the extent of migration and/or population admixture, and as a means to assess effective population sizes and spatial structuring for stock assessment (but see Chapter One and references therein.).

Microsatellites evolve due to mistakes in the replication of the repeat elements. Because of the contiguous nature of the repeats, the replication machinery, at rates potentially as high as once in a thousand replications (Ellegren 2004), often slip resulting in daughter strands that are either missing or gaining a repeat. If back and parallel mutation were not an issue, the divergence of microsatellites would be linear and straightforward. However, this ability to lose or gain iterated elements means that microsatellites possess an analogy of the 'multiple hits' problem of nucleotide saturation in highly mutable DNA sequences (see below). Two individuals with the same sized allele may seem *prima facie* to be identical for that locus, but one individual may have, in its coalescent ancestry, lost a repeat before gaining it again from a second, independent mutation event to return to a state that is the same as that allele of the second individual: that is, microsatellites may exhibit a degree of homoplasy. That alleles can be identical by descent (homologous) or identical by state (homoplasious) is a complexity of microsatellite evolution that is accommodated for by mathematical models. *De facto* population genetic models often assume that each allele is the result of a unique mutation and that all alleles are simultaneously IBD and IBS: the infinite alleles model (IAM, Kimura and Crow 1964). Thus, the IAM may not be suitable for analyzing microsatellite evolution over time frames in which mutation has become an important determining factor in the creation of genetic diversity. An alternate model was created with microsatellite evolution in mind: the SMM (stepwise mutation model (Kimura and Ohta 1978)). This model is the most frequently used model of microsatellite

evolution due to its relative simplicity compared to others that are more complex. Each mutation proceeds with an equal probability of occurring in both the upstream and downstream direction, thus allying greater allelic distance (in terms of base pair separation) with reduced relatedness (Balloux and Lugon-Moulin 2002). Only single repeat units are added or deleted, a fact at variance with observed microsatellite mutation in some taxa (e.g. carp (Yue et al. 2006)). A variant of the SMM model is the Two-Phase Model (TPM) (Valdés et al. 1993) which incorporates this uncertainty in the number of repeats involved in single mutation events (Balloux and Lugon-Moulin 2002). Other models exist which are discussed in numerous reviews (e.g. Balloux and Lugon-Moulin 2002).

As with most tools in the molecular ecologist's toolbox, caution must be exercised by choosing the correct tool for its proper purpose. The high variability of microsatellites is their chief utility and is particularly useful for investigating population structuring. However, for the most part of their history, microsatellite loci have been developed using techniques that seek out the lengthier arrays that are usually more variable than shorter loci (Weber 1990; Ellegren 2000b). This ascertainment bias of enrichment screening (see Zane et al. (2002) for microsatellite development methods) may be misleading, as these loci may not be representative of the genome as a whole (Väli et al. 2008).

2.3.2 The Mitochondrial Genome as Molecular Marker

Mitochondrial DNA markers are commonly used to infer deep historical relationships among populations, to determine and attach dates to lineage splits (Moritz et al. 1987). They have been used in a multitude of intraspecific studies (Avise 2004). Because they are inherited as a single haploid 'type', and because they are also inherited uniparentally - through the maternal line – mitochondrial genomes have an effective size one quarter that of nuclear genomes (in diploid organisms). Because of this reduced effective size relative to nuclear loci, mtDNA is often a more appropriate marker than nuclear loci for identifying higher-level systematic populations (clades, sub-clades, etc), as they are more susceptible to fixation of single-types, or monophyletic lineages, during periods of population fragmentation and isolation or through population bottlenecks. Therefore mtDNA markers provide an excellent starting point to conduct a broad-scale analysis of genetic diversity within candidate species (Avise 1987; Moritz et al. 1987).

The following caveats on the use of mtDNA as a biological marker should be acknowledged: Firstly, these markers represent a matrilineage that may not only reflect population structure as a whole, but may on occasion, where female philopatry exists, reflect the behavioural ecology of the females, and may not take into account male dispersal among demes. Moreover, whilst mtDNA genes have a high rate of mutation compared to nuclear genes (around an order of magnitude greater (Wallace et al. 1994)),[13] mutation rates differ among mtDNA genes. The mitochondrial genome offers a number of structural genes (that encode proteins and ribosomal RNAs) open to genealogical analysis. Amongst the most useful loci for intraspecific studies is the D-loop (see Fig 2.3). Some genes mutate faster than others in many taxa (e.g. the D-loop of the control region) and offer a higher power of resolution than other such loci (e.g. the slowly-evolving genes of the ribosomal RNA). The difference in resolution allows reconstruction of genealogical relationships among haplotypes, populations, species and higher taxonomic categories at different temporal scales. Each temporal scale may be better suited by the application of a differently evolving locus. High mutability of some loci may be detrimental to genealogical reconstruction. A gene that mutates too rapidly will lose phylogenetic signal among haplotypes due to multiple nucleotide 'hits' at particular sites.[14]

The general utility of using the mitochondrial genome as a genetic marker is linked to the relative ease by which it can be assayed within the laboratory, and because of the assumptions of mitochondrial inheritance and biology described above.

[13] Due to the lack of protective histone proteins, close proximity to mutagenic free-radical species and inefficient DNA repair mechanisms (Wallace 1992).

[14] This is because there are only four bases in the genetic code. As time elapses, each base position mutates at a certain rate to one of the other three bases. Eventually, given enough time, and depending upon mutation rate, each position would start to mutate to a third base (including the original base), thus eroding from record the previous mutation and perhaps indicating no evolution at the site in question (even though it would have experienced two mutational 'hits'). Certain mutable 'hotspots' are prone to multiple hits in evolutionary short periods of time. This erosion of genealogical information may compromise studies in which taxon divergence is high and in which the phylogeny is deep.

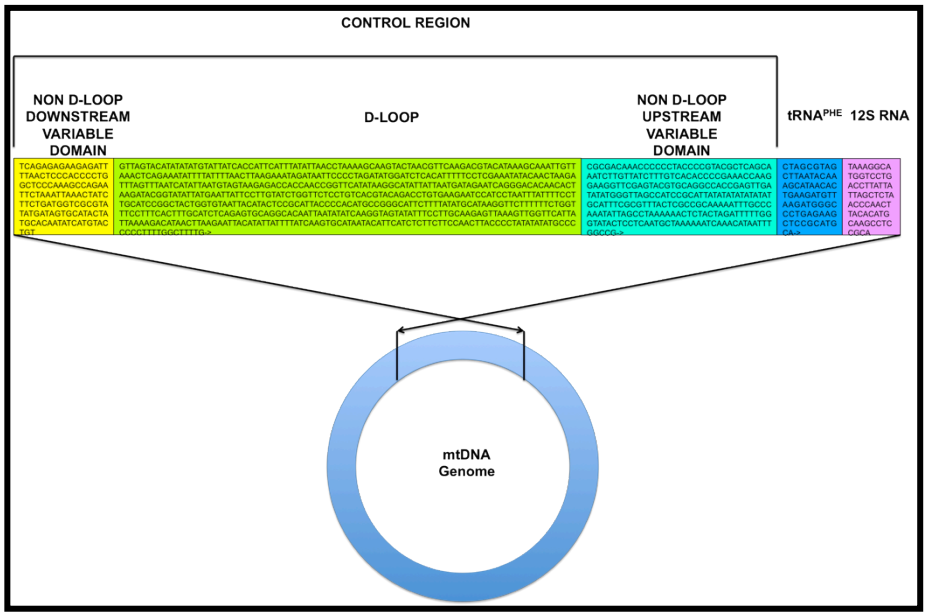


Figure 2.3: The mitochondrial D-Loop as embedded within the control region (CR) of the roach mitochondrial genome15. Shown is the complete control region of a common roach (full sequence derived from GenBank, Accession No: FJ188383.1) with additional upstream sequences that encode the transfer RNA for phenylalanine (blue) and the 5' end of the 12S subunit ribosomal RNA gene (magenta). The 634bp D-loop sequence utilized in this study is highlighted in green. The "D-Loop" sequence is comprised of a central conserved region and variable domains at its upstream and downstream ends (the variable regions located either side of the D-Loop and not sequenced in this study are shaded yellow (downstream – 5') and aquamarine (upstream – 3' end).

Homoplasmy – whereby all mtDNA genomes within a cell are identical - was thought to be prevalent through the action of vegetative segregation, mitochondrial bottlenecks in oogenesis (Bergstrom and Pritchard 1998) and from the prevention of the transmission of paternal mitochondria into the zygote (Birky 1995). However, evidence is accumulating that homoplasmy may not be the norm and that multiple mitochondrial lineages (heteroplasmy) may be more commonplace than initially thought (White et al. 2008). Paternal leakage has been identified in the anchovy (Magoulas and Zouros 1993) and Chinook salmon (Wolff et al. 2008), and recombinant lineages detected in halibut and flatfish (Mjelle et al. 2008 and Hoarau et al. 2002, respectively). Heteroplasmic mtDNA lineages may also occur as discrete size-variants, for example in the sturgeon genera *Acipenser*, *Huso* and *Scaphirhyncus* heteroplasmic variants in the number of iterated repeat units within the 3' end of the control region are found in most species (Ludwig et al. 2000). Thus, the potential exists for heteroplasmic lineages to confuse coalescent-based models and to introduce significant error into estimates of lineage bifurcations in situations where a molecular clock is assumed (Eyre-Walker 2000**;** Slate and Gemmell 2004).

[15] The length of the roach CR is around 1kb (this individual has 999bp) and is somewhat longer than many observed cyprinid CRs (by ~60bp compared to red crucian carp *Carassius auratus*, red variety and blunt snout bream *Megalobrama amblycephala* (Yan et al. 2010), and ~50 bp as compared with the zebrafish *Danio rerio* (Broughton et al. 2001)), although all variable/conserved sequence-block structures are similar to most studied vertebrate taxa (e.g. Roe et al. 1985; Foran et al. 1988).

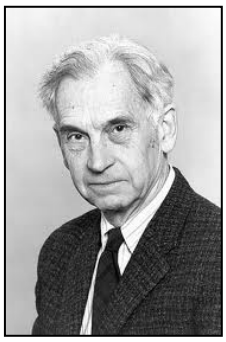
Haplotype networks (unrooted phylogenetic trees based upon shared sequence similarity) are often not fully resolved (a haplotype may be equally inferred to be genealogically linked with two other haplotypes due to a single nucleotide difference separating it from the two putative close relatives), especially in populations that show low sequence divergence. This low resolution in many taxa is either ascribed to low mutation rates or due to recent population bottlenecks. An alternative explanation is recurrent mutation, which may occur in certain hotspots of the mitochondrial control region, causing the problem of multiple hits (Galtier et al. 2006). Additionally, White et al (2008) point out that if heteroplasmy is commonplace, recombination among different mitochondrial lineages will mimic the action of recurrent mutation. This would explain why some taxa are more prone to lineage sorting than others if levels of heteroplasmy, and recombination among heteroplasmic lineages, are variable across taxa. Additionally, the coding mitochondrial genome experiences strong selective pressure to maintain or increase functionality (resulting in stasis or rapid selective sweeps), thus either reducing diversity or showing genetic patterns which may conflate with the expectations of neutral mutations in a population undergoing rapid demographic expansion. Even the non-coding D-loop may be subject to purifying selection (e.g. in humans some variants are more common in individuals who develop certain cancers (e.g. Miyazono et al. 2002; Sharma et al. 2005; and Guo and Guo 2006)). The universality of the above concerns are as yet unknown and the scale of the potential problem(s) open to speculation, but investigators using mitochondrial markers should be cognizant of them and proceed cautiously.

2.4 Epilogue

The choice of genetic marker in molecular ecological studies is primarily contingent upon the questions being asked by the investigator. This chapter has described the main theoretical and methodological theses that underpin much of the science of molecular ecology, although the field is ever-expanding (e.g. adopting highly informative SNP markers and genome-wide screening of genetic variation and adaptive genomics). However, critical assumptions are made regarding the utility of a particular marker to questions pertaining to particular aspects of a species' history or ecology. These assumptions and simplifications, whilst necessary for practical tractability, should accurately reflect the best possible approximation of reality.

It is with this caution in mind, that the study of the genetic diversity of roach was undertaken. The following chapter (Chapter Three) describes the findings of a study that utilises the displacement loop (D-loop) region of the mitochondrial genome to assess the evidence for broad-scale genetic diversity across the natural distribution of this species in the UK; and determines what this biodiversity signifies for the application of fisheries management status, the relationship of UK roach with conspecifics in Europe and beyond and for the postglacial phylogeography of the UK and the European ichthyofauna.

Chapter Three – Mitochondrial DNA Variation Within and Among UK Populations of the Common Roach



"*A species consists of a group of populations which replace each other geographically or ecologically and of which the neighboring ones integrate or hybridize wherever they are in contact or which are capable of doing so (with one or more of the populations) in those cases where contact is prevented by geographical or ecological barriers*." – **early evocation of the Biological Species Concept (BSC), Ernst Mayr (1940).**

3.1 General Introduction

Populations of potamodromous fishes are physically demarcated from similar populations by the geographical location of the bodies of water in which they are found. The ancestors of these fishes, however, may have co-mingled in once shared river networks before the sundering of prehistoric river connections by the dramatic climatic and geological events of past epochs. All sundered populations contain within them gene lineages that may span innumerable generations until they coalesce with homologous sequences found within conspecific individuals located in the same population or in other spatially disjunct populations. Dependent upon the extent of genetic variation in an ancestral population, the historical chronology and geography of a population vicariance and/or dispersal event(s), the mutation rate of the loci under scrutiny and their mode of inheritance, strongly definable population structuring of gene lineages may be uncovered that encompasses a significant geographical component. The potential application of spatial genetic structuring information to evolutionary and ecology theory, systematics, conservation and population management is manifold.

Avise et al. (1979b) first published an unequivocal association between genetic lineages and geography in the pocket gopher *Geomys pinetis*, and soon followed this study with similar studies of geographic patterning of mtDNA within intraspecific populations of birds, fishes and reptiles. The neophyte subject of phylogeography was born, unifying

systematics, palaeontology, population biology and population genetics to describe and explain the spatial patterning of genetic lineages (Avise 2000). The codification of phylogeography as an endeavor incorporating multiple distinct but related specialties provided the basic framework for the application of this new explanatory paradigm within the fields of conservation genetics and population management of wild populations (Avise 1995). Soon, historical biogeography fell under the purview of phylogeography, opening up an almost limitless repository of genetic data from which past demographic inferences could be made and potential causative relationships sought with known or hypothesized geological phenomena.

Perhaps the greatest geological influence upon the distribution of extant organisms in the higher latitudes over the last 2.6 million years has been the periodic glaciations that characterize the Pleistocene epoch. On an almost clockwork basis, the earth has been subject to severe oscillations in climate (Milankovitch cycles, Hays et al. 1976), and its surface has been inundated with a barrage of expanding and contracting ice sheets emanating from both poles. In the face of these climatic intrusions, organisms either had to modify their ranges (by moving away from the ice sheets to lower latitudes, or moving to lower altitudes away from Alpine glaciers) or else go extinct (Berg et al. 2010). Many species with ranges at higher latitude within this period have experienced pronounced demographic contraction and re-expansion into previously glaciated and un-glaciated areas. Depending upon the number and physical connectivity of refugial populations and the ability to disperse effectively in fluctuating climes and habitats and expansion/contact zones, the signal of past history upon current genetic diversity may be inferred. For freshwater fish, routes of ingress and egress from refugia is/was highly dependent on spatially limited routes of dispersal, i.e. river systems. Therefore one may expect a strong signal of both past population fragmentation, but also of current admixture - when two previously separate lineages come into secondary contact – with an increased potential for inferring routes of post-glacial dispersion relative to other, more unrestricted taxa.

The mitochondrial genome is the chief utilitarian marker with which to seek population structuring within and among closely related species (Avise et al. 1987, but see Chapter Two for more details). The application of mitochondrial sequence variation to uncover the extent of genetic structuring within the freshwater fish fauna of northern and central Europe has largely focused upon salmonids (most notably the Atlantic salmon *Salmo salar* (Bermingham et al. 1991; Tessier et al. 1995, 1997; Nielsen et al. 1996; Verspoor 1997; Nilsson et al. 2001; and Tonteri et al. 2005) and the brown trout *Salmo trutta* (Bernatchez

et al. 1992; Osinov & Bernatchez 1996; Weiss et al. 2000; Aurelle & Berrebi 2001; Bernatchez 2001; and Suárez et al. 2001), the grayling *Thymallus thymallus* (Koskinen et al. 2000, 2002; Weiss et al. 2002; Gum et al. 2005, 2009; and Dawnay et al. 2011), the endangered cyprinids of the Iberian peninsula and peri-Mediterranean area (e.g. Mesquita et al. 2001; Salzburger et al. 2003; Mesquita et al. 2005; and Sousa et al. 2008), and those species whose ecology and demography is more suited to inferential studies based upon strong genetic signals resulting from limited inter-population gene flow (e.g. demersal freshwater fishes such as the loaches and sculpins (see Engelbrecht et al. 2000; Kontula & Väinölä 2001; Knapen et al. 2003; Culling et al. 2006; Bohlen et al. 2007; and Šedivá et al. 2008)). Comparatively less attention has been afforded the common and more vagile coarse fishes of Europe, as the human-mediated mixing of lineages from centuries of translocation and stocking are thought to have eroded much of the genetic integrity of isolated river systems.

The majority of these genetic studies of fish utilize the cytochrome b locus within the mitochondrial genome, a marker that is suitable to studies of both intra- and interspecific phylogenetic and phylogeographic relationships. Additionally, many of these studies are inherently interested in the genealogical and demographic processes that contribute to the distribution of extant mtDNA variation, rather than identifying differentiation among populations at mid-high geographic resolutions as a meaningful endeavor in itself. These two avenues of investigation are by no means mutually exclusive, for instance, the phylogenetic relationships among populations and lineages provides the basis for one operational classification of species (PSC – phylogenetic species concept (Nixon & Wheeler 1990)) and comprises a number of definitions for evolutionary significant units, an increasingly important concept in conservation genetics and population management. However, by focusing on the big picture, important details may be overlooked that may prove informative in ascribing management status to individual populations or clusters of populations.

Moritz (1994) states that any form of population unit apt for conservation or management should be based upon diverse and complementary estimates of a putative population's demographic independence, including the use of independently inherited genetic markers. However, there is always an initial step in the management process whereby a population or species of interest needs to be assayed for the distribution of genetic diversity, particularly if no *a priori* knowledge of population structure is known. This initial survey can then be used to guide subsequent, more detailed genetic analyses and/or complement

further non-genetic, more traditional methods of inferring demographic structuring. The adoption of single-locus mtDNA loci as a screening tool is an understandably attractive option for both academic researchers and population managers alike.

The utilization of mitochondrial sequence data, including sequences of the D-loop (or the Control Region) in particular, has been widely adopted for a number of freshwater fish species to investigate population structuring on a *prima facie* basis (e.g. pupfish *Cyprinodon macularius* (Eschelle et al. 2000); bullhead *Cottus gobio* (Knapen et al. 2003); longsnout catfish *Leiocassis longirostris* (Wang et al. 2006); pacu *Piaractus mesopotamicus* (Iervolino et al. 2010); and the goldfish *Carassius auratus* (Takada et al. 2010)). Whilst the D-loop is often considered unsuitable for studying higher-level taxonomic inter-relationships (Brown 1985; Lee et al. 1995), its rapid evolution makes for a potentially statistically powerful genetic marker within species that possess shallow intraspecific phylogenies. Although within the cyprinid family of fishes the D-loop exhibits lower mutation rates compared to other vertebrates (Brown et al. 1986; Shedlock et al. 1992)[16], it is still an optimal single-locus marker to screen for variation. Additionally, D-loop mutation rate variation does exist among cyprinids, probably the result of modulated modes of mitochondrial replication among species (Bielawski & Gold, 2002). Even so, the D-loop may have a mutation rate an order of magnitude greater than that found for a typical nuclear locus (Vigilant et al. 1991), thus underscoring its potential as a marker of spatial patterning of genetic variation.

Whilst it may seem prosaic to determine levels of differentiation, the recording of within-population genetic diversity may be informative from a biogeographic point of view without express reference to a robust phylogeny, which is unavailable for the species-wide distribution of the roach, at least as regards D-loop sequences. Levels of D-loop and, to a lesser extent, cytochrome b sequence differentiation among populations and sequence divergence amongst haplotypes were recorded to survey for the presence of phylogroups (groups that are typified by the presence of rare or atypically frequent haplotype(s)) that may describe acceptable baseline levels of differentiation for further study). This section also addresses the question of whether single drainages merit individual management, as they are *de facto* independent populations by merit of being physically isolated with no presumed natural migration among them (see 'fishery stock' concept of Smith et al. 1990).

[16] Heterothermy, as opposed to homothermy in warm-blooded animals like mammals, was hypothesized by Rand (1994) to account for such inter-taxon differences.

It is expected that the D-loop would reveal more sequence variation than cytochrome b, although overall levels of genetic diversity are expected to be low given the recent glacial history of the British Isles. The latter prediction is consonant with the natural history of higher latitude flora and fauna, and from numerous studies of freshwater fishes (e.g. see Almada & Sousa-Santos (2010) for an in-depth study of environmental factors and historical geography upon mitochondrial and nuclear DNA diversity in fishes of the *Squalius* genus). Moreover, extant populations of organisms are expected bear the genetic imprint of past demographic and non-equilibrium processes. By employing sophisticated analyses of intraspecific genealogies, one may test the hypothesis that UK roach populations, or at least the lineage(s) to which they belong, have undergone a geologically recent population expansion, thereby underscoring Bernatchez & Wilson's (1998) assertion that the genetic diversity of the Holarctic piscifauna has largely been influenced by Quaternary range contractions and expansions.

3.2 Materials and Methods.

3.2.1 Sampling

All sampling efforts were conducted by trained fisheries staff belonging to the UK's Environment Agency, who, as part of their annual surveying of the waterways of England and Wales, sampled each of the chosen locales within England and Wales in this study. EA personnel utilize an electrofishing protocol to stun fish *in situ*. Cycloid scales are then removed from the flanks of each fish for cohort and, in this case, genetic analysis. The scales were immediately stored in paper envelopes, such that any residual moisture was removed (to prevent degradation of DNA), and were then kept in a cool, dry storage space until processed further. Only *Rutilus rutilus sensu stricto* was sampled, and any phenotypically obvious F1 hybrids were discarded. F2 hybrids of roach and common bream *Abramis brama*, although possible to generate in controlled laboratory settings, have not yet been found in the wild (Pitts et al. 1997), indicating that the risk of genetic introgression among these species should pose minimal threat to the interpretation of results.

3.2.1.1 D-Loop

Roach from fifteen hydrologically independent drainages were sampled for variation within a 634bp fragment of the D-loop (see Fig. 2.3, Chapter Two). The fifteen drainages cover the entire natural range of the roach within the UK, located in central, southern, eastern and northeastern England (see Fig 3.1). The river set includes diverse hydrologies inclusive of large, sprawling catchments (e.g. Thames) to small, coastal chalk streams (e.g. Sussex Ouse and Hampshire Avon). Roach from 26 locations were sampled, with multiple sites located within the Thames (main River Thames and four of its tributaries: the Rivers Ash, Kennet, Wey and Thame), the Great Ouse (one upstream site at Newport Pagnell and one downstream at St. Ives, Cambridgeshire), the Trent (three of its tributaries: the Rivers Derwent, Sence and Mease), the Suffolk Stour (upstream site at Stoke-by-Clare and one downstream site at Higham within the main channel), the Yorkshire Ouse (at Beningborough within the main river channel and one site within the River Ure), and at upstream and downstream sites in the Medway and the Yare, respectively. An average of 22 fish were sampled per location (564 individuals/26 sample sites). All pertinent geographical and sampling information can be found in Table 3.1.

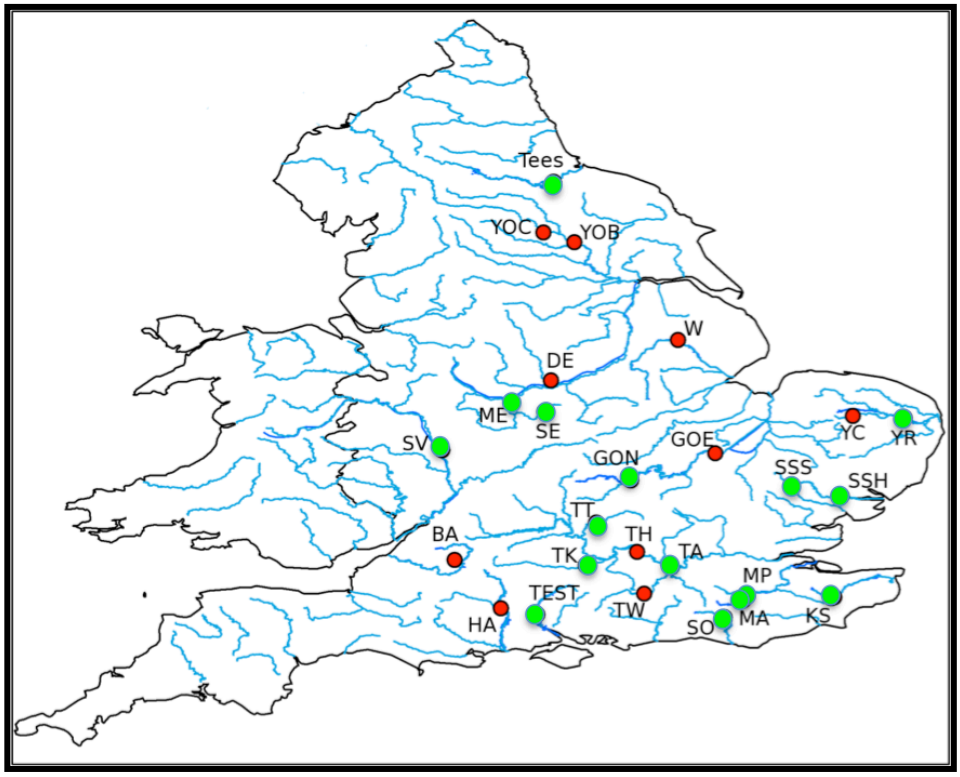


Figure 3.1: Map of sample sites surveyed for mtDNA variation within the UK population of roach (see Table 3.1 for location codes). Red circles denote sites where only D-loop sequences were amplified, whereas both D-loop and cytb sequences were derived from roach at sites denoted by a green circle.

Table 3.1: Sampling information for roach individuals derived from 26 sampled sites across 15 river systems. U = unknown or unrecorded; N = number of individuals sampled at each site for each mtDNA locus (D-loop|cytochrome b).

Drainage	Sampling location	Co-ordinates		Date	N	Code
		Latitude	Longitude			
Thames	Boulter's Reach (Thames)	51.521391	-0.702074	U	40|0	TH
	Lower Benyons (Kennet)	51.412173	-1.114225	27/10/2005	23|10	TK
	Shabbington (Thame)	51.756802	-1.032162	08/09/2005	8|12	TT
	Godalming (Wey)	51.185780	-0.646960	U	30|0	TW
	Gaston Bridge (Ash)	51.398640	-0.431443	U	20|10	TA
Trent	Borrowash (Derwent)	52.901912	-1.397853	26/09/2004	15|0	DE
	Congerstone (Sence)	52.651920	-1.447914	15/05/2006	11|10	SE
	Croxhall (Mease)	52.722099	-1.714048	18/05/2006	9|3	ME
Great Ouse	Newport Pagnell	52.098527	-0.751787	01/05/2005	30|6	GON
	Ely	52.315065	-0.072317	U	23|0	GOE
Yorks Ouse	Beningborough	54.013602	-1.207954	19/07/2004	18|0	YOB
	Westwick (Ure)	54.094084	-1.458707	17/07/2006	17|0	YOU
Yare	Rockland	52.601403	1.4525940	30/08/2006	27|3	YR
	Coston	52.615496	1.0370280	U	6|0	YC
Suffolk Stour	Stoke-by-Clare	52.062391	0.5441150	05/10/2006	26|11	SSS
	Higham	51.971203	0.9451900	U	18|6	SSL
Medway	Penshurst	51.171943	0.1862830	U	6|5	MP
	Ashurst	51.130479	0.1511360	U	5|2	MA
Bristol Avon	Chippenham	51.455021	-2.117534	U	24|0	BAC
Hamps Avon	Salisbury	51.447636	-0.678522	18/05/2004	26|0	HAA
Kent Stour	Ashford	51.156519	0.8853900	04/06/2004	50|6	KSA
Severn	Stourport	52.341262	-2.272966	03/06/2004	37|3	SS
Sussex Ouse	Sheffield Park	50.994653	0.0007130	20/08/2002	13|5	SOS
Tees	Aislaby	54.503500	-1.374557	12/09/2001	40|2	Tees
Test	Timsbury	51.016757	-1.509017	U	7|4	TEST
Witham	Stainfield	53.228549	-0.374611	26/07/2006	35|0	W

3.2.1.2 Cytochrome B

A total of 98 individual roach were sequenced for a 425bp fragment of the cytochrome b locus (Table 3.1 and Fig 3.1). All samples were derived from 16 sites across nine physically isolated drainages, including the Rivers Tees, Kent Stour, Yare, Sussex Ouse and Test. Multiple sites were sampled within the Thames drainage, including sites in the Rivers Ash, Kennet and main River Thames, within the Mease & Sence tributaries of the River Trent and in the Suffolk Stour and River Medway (as for the D-loop).

3.2.2. Laboratory Methods

DNA extraction was based on the protocol of Winnepennickx et al. (1993). Approximately 0.5 - 1cm^2 of roach cycloid scale was finely diced and placed in an Eppendorf tube. The fragmented scales were subsequently immersed in a 300ml solution of cetyltrimethyl ammonium bromide (CTAB) into which proteinase K (10µl of a 10mM solution) was added. Digestion of the tissue and protein complexes was aided by heating the reaction mixture at 55^{o}c overnight. 500mls of chloroform/isoamyl alcohol solution was added and the contents were mixed thoroughly for 10 minutes. Following 10 minutes of centrifugation at 13000 rpm, the supernatant (containing the liberated DNA) was removed and placed in a separate Eppendorf tube into which 500mls of 100% ethanol was added. After a brief period of over-end mixing, the tubes were placed in a -20^{o}c freezer for 45 minutes to aid precipitation of the genetic material. Following this step, the DNA solutions were centrifuged at 13000 rpm for 5 minutes to pellet the DNA. A final round of washing with 70% ethanol (500mls) was carried out, followed by 5 minutes of centrifugation (at 13000 rpm) to pellet the final aggregation of DNA. Excess ethanol was siphoned off and the DNA re-suspended in 100mls of double-distilled water. Each sample yielded more than 50ng µl^{-1} of DNA (after checking for DNA concentration of the final solution on a 1.5 % agarose gel).

3.2.2.1 D-Loop PCR

D-loop sequences were amplified from total DNA via PCR (Saiki et al. 1988). The PCR reagent concentrations and the sequencing reaction volumes for D-loop amplification are as follows: In a total volume of 25µl: 5 µl of 1:10 diluted DNA (around 5ng), 1 µl of each primer (1mM), 0.5 µl of $MgCl_2$ (1mM), 2 µl of 10X buffer, 4 µl of dNTP mixtures (1mM of each dinucleotide) and 0.1µl of *Taq* polymerase (5U µl^{-1} Bioline). The remainder of the reaction volume was double-distilled water (ddH_20). The oligonucleotide primers published in Gilles et al. (2001) were initially used to produce amplicons of the partial D-loop fragment. However, modification of the template thermocycler conditions did not yield a sufficient increase in amplicon concentration, so new primers were designed (using the software Primer 3.0 (Rozen & Skaletsky 2000)).

Forward primer: Roach CRF 5'-TTCTGATGGTCGCGTATATGA-3';
Reverse primer: Roach CRR 5'-TCGGGGTTTGACAAGGATAA-3'.

The thermocycler conditions for these primers are as follows: 14 cycles of denaturation at 92°C for 30 seconds, annealing at 56°C for 45 seconds, and extension at 72°C for 40 seconds; then 20 cycles of 30 seconds each at 90, 56 and 72°C for the denaturation, annealing and extensions steps, respectively. The reaction was terminated after an extra minute at 72°C.

3.2.2.2 Cytochrome B PCR

Partial cytochrome b fragments (425bp) were amplified using the PCR primers published in Gilles et al. (2001), adopting the following thermocycler conditions: an initial 14 cycles consisted of a denaturation step of 30 seconds at 92°C, followed by an annealing step at 48°C for 45 seconds, and then by an extension step of 40 seconds at 72°C. A further 19 cycles were applied of 30 seconds at 90°C, 30 seconds at 48°C and 30 seconds at 72°C for the denaturation, annealing and extension steps, respectively. The reaction was terminated after another 72°C stage at one minute in length. Reaction volumes and reagent proportions for the PCR-mix were identical to those adopted for the D-loop PCR.

3.2.2.3 PCR-Product Processing and DNA Sequencing

Amplified products were cleansed using the 'exosap' procedure. 0.5µl each of exonuclease I and shrimp alkaline phosphatase (SAP) were added to each PCR and heated for 30mins at 35°C, followed by 15 minutes at 85°C (enzymatic denaturing step). The amplified DNA was re-suspended in 25µl of ddH_20. The products were screened on 1% agarose gels to check for clean amplification. Concentrations of over 40ng μl^{-1} were sent for sequencing to Macrogen, Inc. Each completed sequence was checked for clarity using the sequence visualization software 4Peaks version 1.7 (Griekspoor & Groothius, www.mekentosj.com). Following this step, all D-loop and cytochrome b fragments were checked for correct alignment in Clustal X version 1.83 (Thompson et al. 1997).

3.2.3 Statistical Analysis

3.2.3.1 mtDNA Diversity and Population Differentiation

MEGA version 5.05 (Tamura et al. 2011) was used to estimate levels of mtDNA variation within each of the 26 individual locations, the 15 rivers systems and across the UK as a whole for each locus where applicable. For each level of analysis, and for each individual locus, both haplotypic diversity (*h*) and nucleotide sequence diversity (Nei's unbiased estimator π (Nei 1987)) was calculated. Jukes-Cantor estimates of nucleotide divergence (the average number of nucleotide differences per site (Jukes and Cantor 1969)) among pairwise comparisons of river populations were undertaken in DNAsp version 5 (Rozas 2009; Librado and Rozas 2009). Estimates of genetic divergence calculated from the proportion of shared nucleotides, and from the distribution of haplotypic frequencies ($F_{ST,}$ Hudson et al. 1992), were also performed in DNAsp. Exact tests were employed to determine whether the differences in haplotypic content between pairwise comparisons of populations deviates from an expectation of random assortment of haplotypes between population pairs. This analysis was conducted in Arlequin version 3.11 (Excoffier et al. 2005). Additionally, an analysis of molecular variance (AMOVA) (Excoffier et al. 1992) was performed in Arlequin to test for hierarchical genetic structuring. All populations were grouped into river of origin (only those river systems with multiple sampling points could be included in this analysis) to uncover the influence of river system designation upon the distribution of extant mtDNA variation. The statistical significance of the AMOVA analyses was tested with 1000 permutations of the datasets. A smaller, concatenated data set consisting of individuals for whom both loci were sequenced was analyzed in an identical manner.

3.2.3.2 Phylogenetic Analysis

Phylogenetic analysis was carried out to further understand the distribution of D-loop diversity within the British Isles. Uncorrected p-distances between all haplotype pairs (where p = the number of sites at which two haplotypes differ divided by all potentially mutable nucleotide positions) were calculated (in Arlequin 3.11). Phylogenetic reconstruction was attempted using two primary methods: maximum likelihood (ML) and maximum parsimony (MP). For ML inference, the D-loop sequences were first subject to

an initial ML analysis in which the hypothetical evolution of the dataset is compared amongst 88 different models of nucleotide sequence substitution. The most suitable model was estimated by first generating phylogenies from the DNA sequence data via a maximum likelihood procedure in the Phymyl module (Guindon and Gascuel 2003) within the program jModelTest (Posada 2008). The model that best fitted the data was selected (Posada and Buckley 2004). The AIC option (Akaike Information Criterion (Akaike 1969, 1973)) was chosen to select the most appropriate evolutionary model. Model parameters, as suggested by JModelTest, were implemented where possible into the ML analysis. The ML analysis was carried out in Mega 5.05, whereas the MP analysis was carried out in phylogenetic package PHYLIP 3.69 (Felsenstein 1989, 1993), with bootstrap re-sampling to gauge confidence on internal nodes within trees. A final bootstrap consensus tree was drawn. In both instances, phylogenies were rooted by the inclusion of D-loop sequences from one individual from the congeneric *Rutilus rubilio* (GenBank accession number: AJ388400.1) and three from the confamilial chub *Squalius cephalus* (accession numbers: AY301921.1; AY301920; AJ388429). Further, the likelihood that the sequence data is supportive of a strict molecular clock was tested upon the ML tree topology using a likelihood test (Tamura 1992), also implemented in MEGA 5.05. Comparing likelihood scores when an assumption of clocklike behavior is withheld and when it is relaxed tests the null hypothesis of clocklike evolution within the inferred phylogeny.

To determine whether there is any broad correspondence between geographical location and phylogenetic relationships amongst haplotypes, a haplotypic network was constructed of all roach D-loop, cytochrome b and concatenated data haplotypes. Firstly, statistical parsimony (Templeton et al. 1992) was implemented in the TCS version 1.18 module (Clement et al. 2000) within the program ANeCA version 1.2 (Panchal et al. 2007). The remaining two methods invoked are: median-joining networks (MJN, Bandelt et al. 1999) and maximum parsimony (MP) phylogenetic reconstruction with tree-conversion (Salzburger et al. 2011). Bandelt et al.'s median-joining algorithm was implemented in the program Network version 4.6 (Flexus Engineering, 2011). When genealogical relationships are assumed to be shallow, some degree of reticulation within the network ("loops") is expected, whereby each of two parsimonious network connections is equally likely (Salzburger et al. 2011). In a study of 1000 known, simulated haplotype networks, Salzburger and colleagues found that in situations in which the genealogy is shallow and migration is likely to be high MP reconstruction outperforms the TCS algorithm. Therefore, an unrooted haplotypic network was constructed that was derived from a

consensus MP tree calculated without outgroups in PHYLIP and converted within the online program Haplotype Viewer (http://www.civat.at/~greg/haploviewer).

To estimate the mutation rate of D-loop in the common roach, from which estimates of lineage bifurcations are based, a Bayesian approach was used to estimate, from the data, the rate of mutation along each lineage within a wider D-loop phylogeny derived from all included sequences. Applying the sequence evolution model suggested by JModelTest in the program BEAST v. 1.7.4 (Drummond et al. 2012), a phylogeny was constructed according to Bayesian principles, and priors were used to date the root of the tree and to model the possible demographic scenarios which may have differential impacts upon temporal spacing of mutation and branching events within the phylogeny. In the first instance, each reconstructed phylogeny was rooted using outgroup sequences utilized for the previous phylogenetic analyses. From fossil evidence, all modern cyprinid lineages in their generic forms appear in the fossil record by the mid-Miocene epoch, some 13.6 million years ago (Cavender 1991). However, the *Squalius* genus appears to have originated in the late Oligocene (23-28 million years ago, de La Peña 1995), along with many unidentifiable forms of Leuciscinae (Bŏhme & Ilg 2003), with the age of the chub *Squalius cephalus* to be at least 6.56 million years old (Bŏhme, unpublished), whilst the earliest known *Rutilus* specimens (*Paleorutilus*) are dated to 14.5 – 15.5 million years ago (mid-Miocene, Schulz-Mirbach & Reichenbacher 2006). Three chub individuals (GenBank accession numbers AY301920, AY301921 and AJ388429) and one *Rutilus rubilio* sequence (Genbank accession number: AJ388400.1) were used as the outgroup sequences.

A hard minimum age of 23Mya was applied to a lognormal distribution of possible coalescent events, from which the roach and the *Squalius* lineages diverged, with a mean in real space of 23Mya and standard deviation of 1.0. This allowed for a soft upper bound on the age of the root allowing for uncertainty associated with the dating of fossil evidence and the actual origination of a taxon in palaeohistory. No fossil calibration was applied to the *Rutilus* genus node, as only a subset of *Rutilus* species have been sequenced for the D-loop. A relaxed clock was incorporated into the evolutionary model (Drummond et al. 2006), incorporating branch lengths that may vary in their lengths due to a combination of time since divergence and mutation rate heterogeneity among and within branches. A lognormal distributed, uncorrelated clock model was implemented, whereby estimates of mean mutation rate are relaxed among lineages. A lognormal prior was applied to the mean rate of nucleotide substitution (mean rate of 0.025 substitutions per site per lineage

per million years (= 5% sequence divergence rate per million years) with a standard deviation of 1, such that the 95% posterior distribution covered a mutation rate embracing a low of 0.002 nucleotide substitutions per site per million years and 0.1 nucleotide substitutions per site per million years)[17].

Because different demographic scenarios can influence the temporal positioning of substitutions within a tree, a number of demographic scenarios were implemented (one speciation model - a Yule speciation model; two coalescent models - constant population size and a Bayesian linear-growth skyline analysis; and a model incorporating a Yule model for internal nodes (such as those leading to the split with *Squalius* and *Rutilus rubilio*, but applying a coalescent model to deal with the author's focus on the common and Caspian roaches, for whom past demographic history may have played a significant role in shaping recent genealogies)), and a Bayes Factor analysis, in which the harmonic mean method of Newton & Rafferty (1994) was applied to compare the marginal likelihoods of each hypothesized model (Suchard et al. 2001), was carried out in the program Tracer v. 1.5 (Rambaut & Drummond 2004) to determine which model best suited the data. From the analysis, the mean mutation rate over all lineages and mean rate among lineages was determined with 95% HPD limits, in addition to the dating of important coalescent events within the roach clade(s). Each BEAST analysis was run for 5 x 10^6 iterations, sampling every 1000^{th} step. The first 5 million steps were discarded as burn-in. Each run was analyzed in Tracer to ascertain whether the output had achieved convergence and adequate mixing of chains (ESS of all parameters >> 200). BEAST was run in conjunction with BEAGLE (Ayres et al. 2012), a suite of applications for running intensive Bayesian computations. All 50% clade credibility trees were created and annotated in TreeAnnotator v 1.6.2 and visualized in FigTree v. 1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/).

[17] It is unwise to expect similar rates of mutation or strict clocklike behavior within broad taxonomic groups. The control region in nine-spined sticklebacks does not adhere to rigid clock expectations (ranging from 0.98% sequence change per million years to 2.91% sequence change per million years across species (Takahashi & Goto 2001)). The control region in the bullhead was found to evolve at a rate approaching 9% divergence per million years (Volckaert et al. 2002). Additionally, in both salmonids and cobitids, the rate of sequence divergence amongst control region sequences was found to be less than that for the cytochrome b locus when comparing closely related species, but not when phylogenetic distances among taxa were greater (Bernatchez and Danzmann 1993; Tang et al. 2006).

3.2.3.3 Demographic Analysis

In addition to assessing the potential for past demographic influence upon extant intraspecific genealogies of roach via the BEAST analyses, a mismatch analysis (Li 1977; Slatkin and Hudson 1991; Rogers and Harpending 1992; and Schneider and Excoffier 1999) was implemented in Arlequin 3.11, in which the observed distribution of pairwise nucleotide differences within the dataset is compared to the expected distribution under a model of population expansion or stasis. Harpending's raggedness index (Harpending 1994) was appropriated to assess the goodness of fit of the observed values to theoretical expectations. Close approximation is determined by bootstrap re-sampling over 10000 replicates and a *p*-value ascertained ($p < 0.05$ = the probability that the iterated simulation ≥ observed raggedness). To determine to what extent the data support the null hypothesis of long-term population stability based upon deviations from neutral expectations, a number of tests were conducted in DNAsp: Tajima's D (Tajima 1989) and Fu's Fs (Fu and Li 1993; Fu 1997) statistic, along with measures of their significance (after 10000 bootstrap re-samplings of the coalescent distribution, made possible in DNAsp, in which the final mutation-scaled population size θ was set at 1000 individuals), were calculated. In addition to the population expansion tests, the R_2 statistic of Ramos-Onsins & Rozas (2002) was implemented as a more powerful method to reject the null hypothesis of a constant population size. Values of R_2 are expected to be lower in populations undergoing range expansion than those in which population size is constant. This significance of the R_2 test was assayed by implementing the coalescent simulation package within DNAsp also used to test the significance of theoretical deviations from neutrality. Following Hänfling et al. (2009), the null hypothesis of a constant population was rejected if the calculated value was found to lie outside of the 95% CI distribution.

To calculate an estimate of the time, *t* (in years), since the UK population experienced demographic expansion, the equation:

$$t = \frac{\tau}{2(k\mu)}$$

(Slatkin and Hudson 1991) was employed (where τ is an expansion term, *k* is the length of nucleotide sequence being compared and μ is the hypothesized nucleotide substitution rate per site per year per lineage).

3.3 Results

3.3.1 mtDNA Diversity

3.3.1.1 D-Loop

A total of 564 individual roach from the UK were typed for nucleotide sequence variation at a 634bp fragment of the D-loop locus within the mitochondrial genome. A survey of the 358.14kbp dataset yielded 18 unique haplotypes based upon variation in nucleotide content at 14 segregating sites. Of the 14 segregating sites, four are singletons, while the remaining ten are parsimony informative. The singleton sites are located 53, 80, 166 and 369bp from the 5' end of the sequence, whereas the parsimony informative sites are located 119, 152, 325, 326, 327, 426, 456, 461, 493 and 520bp from the 5' end of the fragment (see Table 3.2). The average proportion of the four different bases were not equal within the D-loop, with thymine accounting for 35.5% of all nucleotide positions, cytosine 19.4%, adenine 31.9% and guanine 13.2%. These proportions are representative of other cyprinids (e.g. vairone *Leuciscus souffia* (Salzburger et al. 2003)) and teleost control region sequences as a whole (Jean et al. 1995**)**. The relative bias of transitions to transversions was calculated by using a maximum likelihood analysis in MEGA 5.05, assuming Tamura's 3-parameter model (Tamura 1992) that incorporates decoupled mutation rates between transitions and transversions and skewed nucleotide frequencies. The probability of a transition mutation occurring at any one site is 2.37 times more probable than a transversion (R = 2.37, likelihood = -980.238), a finding consistent with the majority of DNA sequences (e.g. Fitch 1967; Gojobori et al. 1982). Similar levels of transitional bias are found under various substitution models for this dataset (e.g. Kimura's 2-parameter model (Kimura 1980): R = 2.26, likelihood = -1019.293; HKY (Hasegawa et al. 1985): R = 2.40, likelihood = -975.640; Tamura & Nei (1993): R = 2.40, likelihood = -974.620; General Time Reversible model (GTR, Nei & Kumar 2000): R = 2.40, likelihood = -973.295), indicating that not enough sequence divergence has accrued to justify the use of more sophisticated evolutionary models of sequence evolution in future analyses of D-loop variation, at least within and among UK haplotypes. Simpler models should be used to lessen the greater variances associated with these more complex models (Nei & Kumar 2000).

Table 3.2: Variable nucleotide positions within the 634bp fragment of the D-loop.

Hap	Nucleotide position													
	53	**80**	**119**	**152**	**166**	**325**	**326**	**327**	**369**	**426**	**456**	**461**	**493**	**520**
D1	A	G	A	C	A	C	T	G	T	A	G	T	A	T
D2	*	*	G	*	*	*	*	*	*	*	*	*	*	*
D3	*	*	*	*	*	*	C	*	*	*	*	*	*	*
D4	*	*	*	*	*	T	*	C	*	*	A	*	*	*
D5	*	*	*	*	*	*	C	*	*	G	*	*	*	*
D6	*	A	*	*	*	*	*	*	*	*	*	*	*	*
D7	*	*	*	*	*	*	*	*	*	*	A	*	*	*
D8	*	*	*	*	*	*	C	*	A	*	*	*	*	*
D9	*	*	*	*	*	*	*	C	*	*	A	*	G	*
D10	*	*	*	*	*	*	*	C	*	*	A	*	*	*
D11	*	*	*	*	*	*	C	*	*	*	*	*	*	C
D12	*	*	*	*	*	*	*	*	*	*	*	C	*	*
D13	G	*	*	*	*	*	*	*	*	*	*	*	*	*
D14	*	*	*	A	*	*	*	*	*	*	*	*	*	*
D15	*	*	G	*	*	*	C	*	*	*	*	*	*	*
D16	*	*	*	*	*	*	*	*	*	*	*	*	*	C
D17	*	*	*	*	T	*	*	*	*	*	*	*	*	*
D18	*	*	*	*	*	*	A	*	*	*	*	*	*	*

* Asterisks indicate that the nucleotide position possesses an identical nucleotide to that of haplotype D1.

Table 3.3 shows the haplotype counts for each sampled location within the UK. Of the 564 individual fish sequenced at this locus, 396 were all identical for the same haplotype (D1, 70.213% of all samples). As can be observed both in Table 3.3 and pictorially on the map in Fig 3.2, haplotype D1 is ubiquitous throughout the sampled range of the roach within the United Kingdom, maintaining majority frequencies in all populations apart from the populations within the River Severn, and at Sheffield Park, Sussex Ouse and Lower Benyons in the River Kennet, Thames drainage. In the Kennet population, D1 is still the most frequent individual haplotype, but assuming that the sampled haplotype frequencies are representative of their real frequencies, then a randomly chosen Kennet fish is likely to possess a haplotype other than D1 (D3, D12, D14 or D15).

In the Severn and Sussex Ouse, however, only two haplotypes are present: Severn: D1 and D3 at frequencies of 21.62% (8/37) and 88.38% (29/37), respectively; Sussex Ouse: D1 and D10 at frequencies of 46.15% (6/13) and 53.85% (7/13), respectively. In the Coston population within the Yare, frequencies of D1 and D3 are 50:50, although the sample size here is only 6 individuals. The second most common haplotype – D3 – is also fairly widespread within the UK as a whole, found in 20 out of 26 (76.92%) sites and in 84 out of the 564 (14.89%) surveyed fish. As noted above, D3 achieves majority frequency status within the Severn population. D3 is particularly prevalent in southern, central and western populations, but less so in the east (with the exception of the upper Yare sample). D7 is the

third most prevalent haplotype (26/564, 4.61%). Apart from three fish in the Tees and three in the Great Ouse, the remaining 20 roach with this haplotype are distributed among drainages that discharge into the English Channel. The final 'common' haplotype (D12, 14/564 = 2.48%) is also found in drainages with a southern distribution and which empty into the English Channel. Half of the haplotypes are found within the Thames (3 individuals in the River Ash, and two individuals each within the main Thames as well as the River Kennet). The remaining seven haplotypes are found in the Bristol Avon (2 roach) and the Hampshire Avon (5 roach).

The remaining haplotypes are rare. Those haplotypes that number more than two tend to cluster within few populations, e.g. D2 is found in 5 individuals in the upper reaches of the Suffolk Stour at Stoke-by-Clare (the remaining haplotype found in the Yorkshire Ouse). D5 is likewise found in a single fish in the Yorkshire Ouse (River Ure tributary), but is also found in the main in the Tees (5 roach). D10 is distributed between the Sussex Ouse (7 roach) and the upper Great Ouse (2 roach); the 6 individuals possessing D11 are scattered among the upper Great Ouse (1 roach), the Main River and Ash tributary of the Thames (3 roach) and the Medway at Ashurst (1 roach). The remaining haplotypes are all found in a single population: D4 (2 roach within the River Ure); D6 (1 roach in the Great Ouse at St. Ives); D8 (1 roach) and D9 (3 roach) in the upper stretch of the Great Ouse; D13 (one individual in the Thame tributary of the Thames); D15 (1 roach in the Kennet); D16 & D17 (one individual each) in the Tees; and D18 is found in a single roach within the River Derwent in the Trent catchment.

The number of distinct haplotypes found in any one sampled population ranged from one (in the Rivers Witham and Suffolk Stour at Higham) to seven in the upper Great Ouse at Newport Pagnell. The mean number of unique haplotypes per population is 3±1.5 (mean and standard deviation). Comparing drainages, the average number of haplotypes per river system is 3.73±2.2 (ranging from the invariant Witham to the 8 haplotypes found within both the Thames and the Great Ouse, see Table 3.4). Of the 19 haplotypes found within the UK, 8 are private haplotypes found in only a single location. The number of unique haplotypes found in each bears some relation to the size of the catchment (Pearson's r = 0.4763, one-tailed p-value = 0.036).

Levels of haplotypic and nucleotide diversity over all 26 samples was 0.482±0.024 and 0.00101±<0.001, respectively (see Table 3.3). Haplotypic diversity ranged from zero in the fixed populations of Witham and the lower Suffolk Stour to a maximum of

0.753±0.046 in the River Kennet sample, followed by 0.733±0.155 at Penshurst in the Medway. Extremely low levels of non-zero haplotype diversity were observed in the Yare at Rockland (0.074±0.067). Because of zero variation in the number of haplotypes in the Witham and lower Stour samples, nucleotide diversity here was also zero. The next lowest nucleotide diversity score was in the Rockland area of the River Yare (π = 0.0025±0.0023), with a similarly low level of diversity in the River Yorkshire Ouse sample (π = 0.0037±0.0032). The maximum degree of nucleotide diversity was observed in the upper reaches of the Great Ouse (π = 0.0405±0.0088), followed by the River Ure (π = 0.0363±0.0013), River Kennet (π = 0.0361±0.0047), and the Sussex Ouse (π = 0.0359±0.0040). Over all sample sites, mean haplotypic and nucleotide diversity was 0.4026±0.1037 and 0.0178±0.0048, respectively, and the average number of nucleotide differences per pairwise comparison was 0.639.

Table 3.3: Haplotype counts and estimates of D-loop diversity for each of 26 sampling sites within the UK.

Pop	Haplotype*																		Σ	Haplotype		Nuc diversity	
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16	D17	D18		*h*	s.d	*π*	s.d.
YOB	17	1																	18	0.111	0.096	0.0037	0.0032
YOC	12		2	2	1														17	0.500	0.135	0.0363	0.0013
GOE	18		3			1	1												23	0.383	0.120	0.0137	0.0047
GON	19		2				2	1	3	2	1								30	0.593	0.100	0.0405	0.0088
TH	28		6				1				3	2							40	0.491	0.087	0.0216	0.0046
TA	15		1								1	3							20	0.432	0.126	0.0186	0.0067
TT	6		1										1						8	0.464	0.200	0.0167	0.0079
TK	8		7									2		5	1				23	0.753	0.046	0.0361	0.0047
TW	27		3																30	0.186	0.088	0.0062	0.0029
SV	8		29																37	0.348	0.078	0.0116	0.0080
BA	21		1									2							24	0.236	0.109	0.0081	0.0039
SSH	18																		18	0	0	0	0
SSS	20	5	1																26	0.385	0.102	0.0133	0.0038
YR	26		1																27	0.074	0.067	0.0025	0.0023
YC	3		3																6	0.600	0.129	0.0200	0.0043
HA	15		5				1					5							26	0.615	0.081	0.0240	0.0043
KS	27		6				17												50	0.590	0.043	0.0225	0.0024
W	35																		35	0	0	0	0
TEES	27		3		5		3									1	1		40	0.529	0.088	0.0256	0.0053
DE	14																	1	15	0.133	0.112	0.0044	0.0037
ME	7		2																9	0.389	0.164	0.0130	0.0055
SE	7		4																11	0.509	0.101	0.0170	0.0034
SO	6									7									13	0.538	0.060	0.0359	0.0040
TEST	5		2																7	0.476	0.171	0.0159	0.0057
MP	3		2				1												6	0.733	0.155	0.0289	0.0082
MA	4										1								5	0.400	0.237	0.0267	0.0158
Σ	396	6	84	2	6	1	26	1	3	9	6	14	1	5	1	1	1	1	564	**0.403****	**0.1037**	**0.0178**	**0.00482**

* Haplotype colours correspond to those of the pie chart segments in Fig 3.2.

** Mean values in bold type.

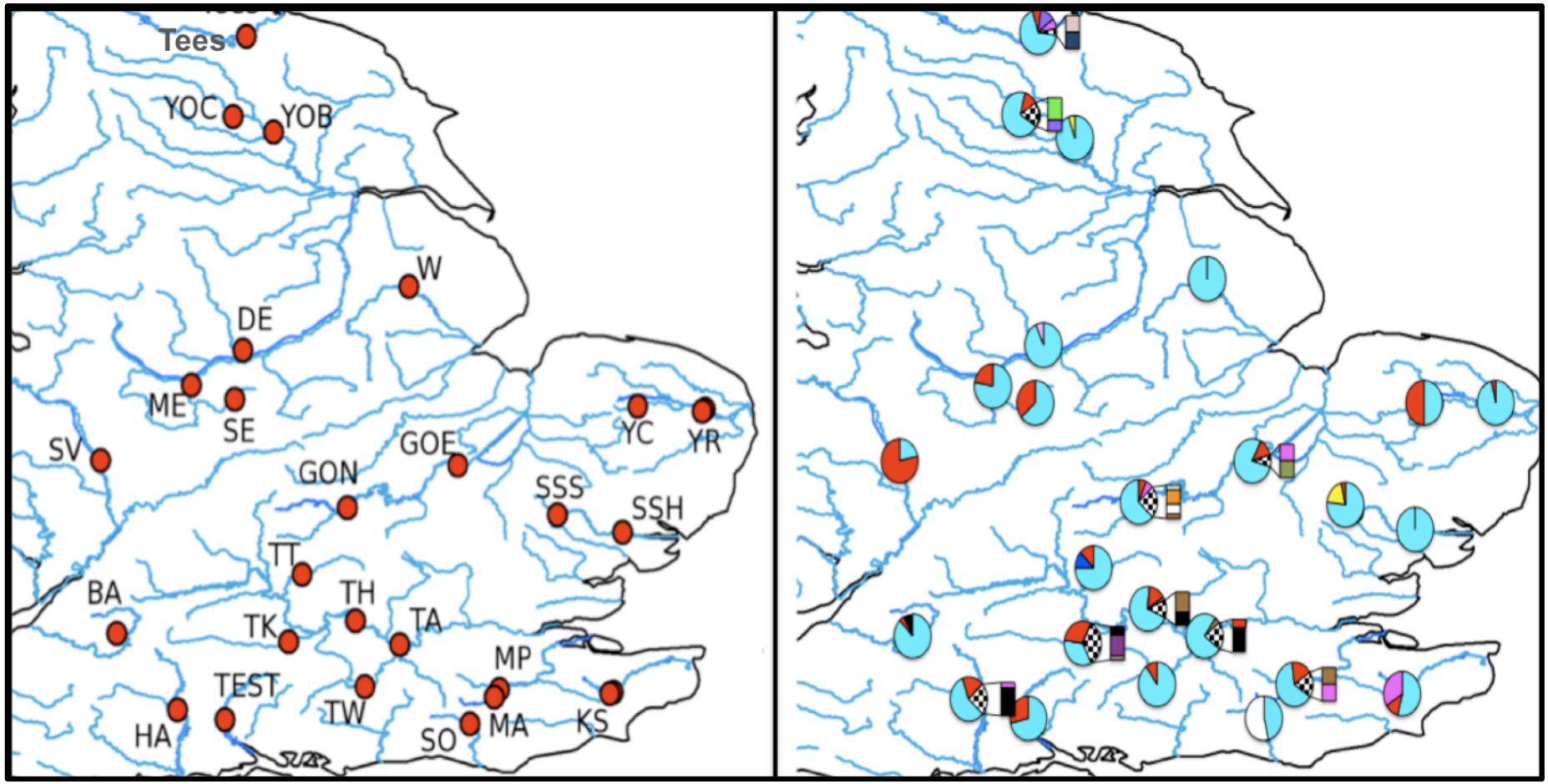


Figure 3.2: Frequency distribution map of D-loop haplotypes. Right: Haplotype colours correspond to those in Table 3.3. Bar charts are shown for low frequency haplotypes when the number of haplotypes per location is > 3. Left: Sample site reference map.

Table 3.4: Diversity data for each catchment listed by decreasing area (N = number of samples; H = number of haplotypes; area is km^2

Catchment	Area	N	H	Hap diversity		Nuc diversity	
				h	s.d	π	s.d.
Thames	13513	121	8	0.4940	0.0510	0.0205	0.0026
Severn	11381	37	2	0.3480	0.0780	0.0116	0.0080
Yorkshire Ouse	10611	35	5	0.3140	0.1000	0.0203	0.0078
Trent	10329	35	3	0.3390	0.0890	0.0113	0.0030
Great Ouse	8443	53	8	0.5040	0.0810	0.0297	0.0063
Yare	3017	33	2	0.2200	0.0870	0.0073	0.0029
Hamps Avon	2994	26	4	0.6150	0.0810	0.0240	0.0043
Witham	2915	35	1	0.0000	0.0000	0.0000	0.0000
Suffolk Stour	2806	44	3	0.2460	0.0800	0.0084	0.0028
Kent Stour	2720	50	3	0.5900	0.0430	0.0225	0.0024
Bristol Avon	2229	24	3	0.2360	0.1090	0.0081	0.0039
Tees	1792	40	6	0.5290	0.0880	0.0256	0.0053
Medway	1608	11	4	0.6000	0.1540	0.0267	0.0086
Test	1206	7	2	0.4760	0.1710	0.0159	0.0057
Sussex Ouse	574	13	2	0.5380	0.0600	0.0359	0.0040
μ	5076	37.6	3.73	**0.4033**	**0.0848**	**0.0179**	**0.0045**

3.3.1.2 Cytochrome B

The survey of partial cytochrome b variation from 98 roach individuals derived from 13 English rivers yielded a total of 6 unique haplotypes. This 425bp fragment of cytochrome b unveiled five polymorphic sites (four singleton sites (at positions: 204, 261, 330 and 334) and one site that is parsimony informative (position: 352)). All polymorphic sites and their position within the amplified fragment of cytochrome b are shown in Table 3.5. The base composition of the cytochrome b fragment, averaged over the 98 individuals, is as follows: T: 28.2%; C: 26.8%; A: 27.8%; and G: 17.2%, indicating guanine deficiencies similar to the levels found in other teleost fishes (e.g. in *Salvelinus* spp., Radchenko 2004), with deficiency greatest at the 2nd and 3rd codon positions and negligible differences at the first position. All mutations within the dataset were transitions (one among-pyrimidine transition (C - >T) and four among-purine transitions (2 each of: A - > G and G -> A), referenced to the base composition of the "NEW" haplotype) (Table 3.5)).

Table 3.5: Variable nucleotide sites within a 425bp fragment of the cytochrome b gene uncovered by a survey of 98 individual roach specimens from among 11 physically distinct drainages.

Hap	Nucleotide Position				
	204	261	330	334	352
NEW	C	G	G	A	G
UK	•	A	•	•	•
UKSS	•	•	•	•	A
UKTh	•	•	•	G	•
UKTe	•	•	A	•	•
KenSo	T	•	•	•	•

Pseudogenes found within the nuclear genome derived from ancestral, functional mtDNA (see review in Benasson et al. 2001) may prove to be a serious source of error in the estimation of relationships among haplotypes and of divergence among populations or lineages (Zhang & Hewitt 1996). All cytochrome b sequences were therefore checked for the presence of mid-sequence stop codons – which tend to accumulate in nuclear pseudogenes (Vanin 1985; Mighell et al. 2000), but would soon be erased from the mitochondrial population through purifying inter-organelle (intracellular) selection (e.g. Walsh 1992; Mamirova et al. 2007) within the main coding body of cytochrome b. All samples conformed to the expectations of a functioning, translational coding sequence.

Tallies of each haplotype found in each sampled location for which cytochrome b sequence data is available is given by Table 3.6 and illustrated graphically in Fig 3.3. With reference to Table 3.6, an overwhelming majority of roach individuals possessed the "NEW" haplotype (N = 89, 90.82% of all sampled fish). Like the D1 haplotype of the D-loop, "NEW" is to be found in every location in which fish were sampled (Fig. 3.3). For 9 out of the 16 sampled locations, it was the only haplotype found (with sampled individuals ranging from 3 – 12). For all other cases, "NEW" was the dominant haplotype with the sole exception of the River Tees from which only two individual fish were sequenced. The only other haplotype to number greater than one is the "UK" haplotype (N = 5, 5.10% of all sampled fish), which is found in one roach from each of the Rivers Kennet, Mease and Sence, and from within two fish from the Sussex Ouse.

Table 3.6: Haplotype counts and estimates of cytochrome b diversity for each of 16 sampling sites within the UK

Pop	Haplotype*						Σ	Hap diversity		Nuc diversity	
	NEW	UK	UKSS	UKTh	UKTe	KenSo		*h*	s.d.	π	s.d.
GON	6						6	0.000	0.000	0.00000	0.00000
TA	9			1			10	0.200	0.154	0.00047	0.00036
TT	12						12	0.000	0.000	0.00000	0.00000
TK	8	1				1	10	0.378	0.181	0.00094	0.00048
SV	3						3	0.000	0.000	0.00000	0.00000
SSH	6						6	0.000	0.000	0.00000	0.00000
SSS	10		1				11	0.182	0.144	0.00043	0.00034
YR	3						3	0.000	0.000	0.00000	0.00000
KS	6						6	0.000	0.000	0.00000	0.00000
TEES	1				1		2	1.000	0.250	0.00240	0.00120
ME	2	1					3	0.667	0.314	0.00157	0.00074
SE	9	1					10	0.200	0.154	0.00047	0.00036
SO	3	2					5	0.600	0.175	0.00140	0.00040
TEST	4						4	0.000	0.000	0.00000	0.00000
MP	5						5	0.000	0.000	0.00000	0.00000
MA	2						2	0.000	0.000	0.00000	0.00000
Σ	89	5	1	1	1	1	98	**0.202****	**0.086**	**0.00050**	**0.00020**

* Haplotype colours correspond to those of the pie chart segments in Fig 3.3.
** Mean values in bold type.

The remaining four haplotypes are each found in a single individual in just one location: "UKSS" (in the upper Suffolk Stour at Stoke-by-Clare), "UKTh" (in the River Ash, Thames catchment), "UKTe" (River Tees) and "KenSo" (in the River Kennet, Thames catchment). The maximum number of distinct haplotypes found within a population was three, in the River Kennet.

Accordingly, levels of nucleotide diversity are low: mean (+ s.d.) nucleotide diversity over the 16 locations = 0.0005±0.0002 (ranging from lows of zero in a majority of locations to a high of 0.0024±0.0012 in the Tees). Mean haplotypic diversity was similarly limited: mean = 0.2017±00858 (ranging from zero to one in the Tees).

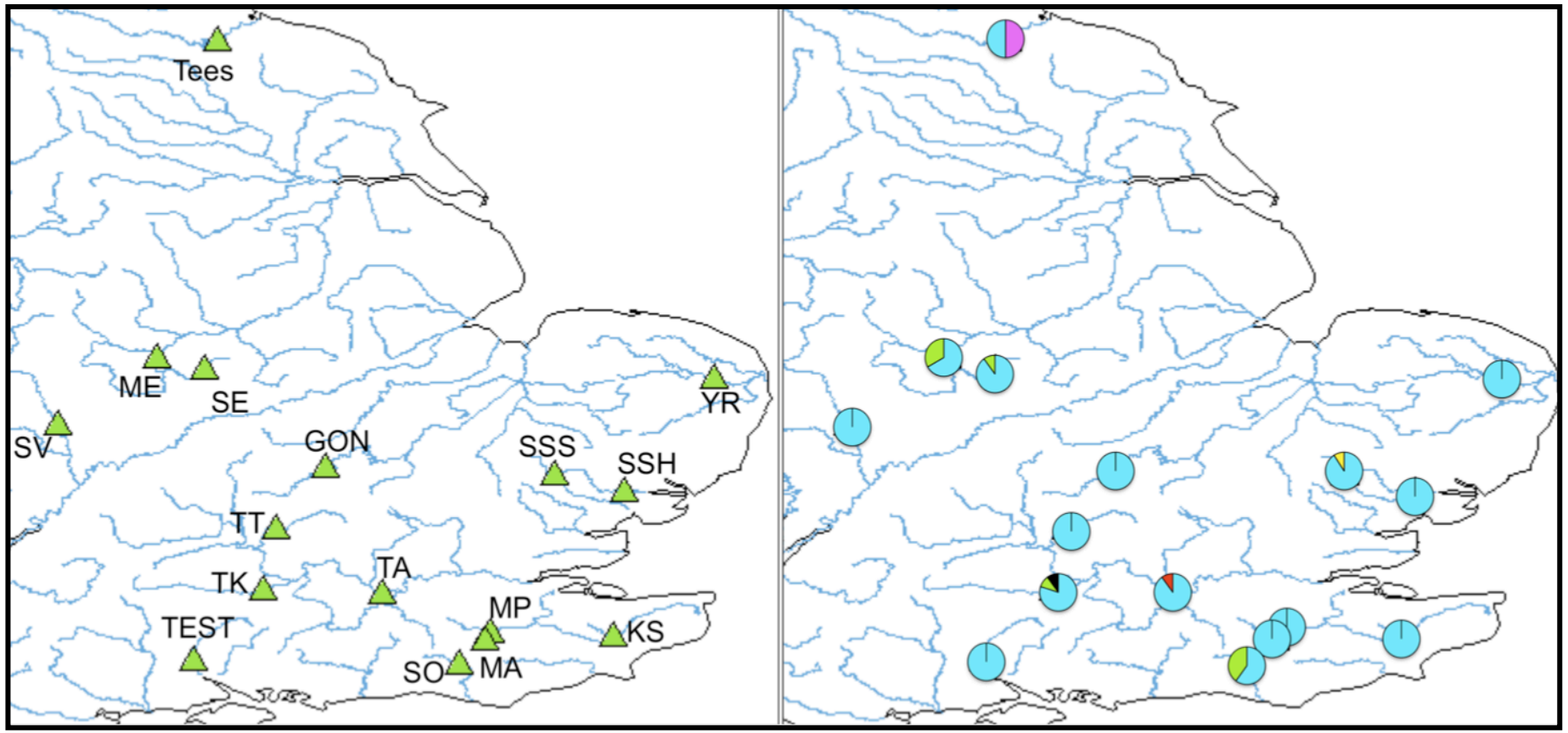


Figure 3.3: A frequency distribution map of cytochrome b haplotypes. Right: Haplotype colours correspond to those in Table 3.6; Left: Sample site reference map.

Table 3.7: Cytochrome b diversity data by catchment (in decreasing order of size). N = number of samples; H = number of haplotypes; area is km^2

Catchment	Area	N	H	Hap diversity		Nucleotide diversity	
				h	s.d	π	s.d.
Thames	13513	32	4	0.1810	0.0900	0.0004	0.0002
Severn	11381	3	1	0.0000	0.0000	0.0000	0.0000
Trent	10329	13	2	0.2820	0.1420	0.0007	0.0003
Great Ouse	8443	6	1	0.0000	0.0000	0.0000	0.0000
Yare	3017	3	1	0.0000	0.0000	0.0000	0.0000
Suffolk Stour	2806	17	2	0.1180	0.1010	0.0003	0.0002
Kent Stour	2720	6	1	0.0000	0.0000	0.0000	0.0000
Tees	1792	2	2	1.0000	0.5000	0.0024	0.0012
Medway	1608	7	1	0.0000	0.0000	0.0000	0.0000
Test	1206	4	1	0.0000	0.0000	0.0000	0.0000
Sussex Ouse	574	5	2	0.6000	0.1750	0.0014	0.0004
μ	5217	8.91	1.64	**0.1983**	**0.0689**	**0.0005**	**0.0002**

Table 3.7 re-groups measures of cytochrome b diversity by river drainage. Mean levels of haplotypic and nucleotide diversity over all catchments are 0.1983±0.0689 and 0.0005±0.0002, respectively. Only five river systems were polymorphic for the cytochrome b locus (Thames, Trent, Suffolk Stour, Tees and Sussex Ouse). The Tees' two different haplotype-bearing individuals maximize haplotypic diversity. Additionally, the Sussex Ouse has high diversity due to the presence of two haplotypes within a small sample of five individuals (h = 0.600±0.175, π = 0.0014±0.004). The Suffolk Stour population possesses relatively little variation given the higher sampling effort in this catchment (N = 17, h = 0.118±0.101, π = 0.0003±0.0002). The Thames catchment was subject to the greatest number of sequenced roach (N = 32), although diversity here was fairly low despite having the largest number of unique haplotypes: four (h = 0.181±0.090, π = 0.0004±0.0002). Unlike the D-loop, there was no indication of a relationship between catchment size and diversity contained therein.

3.3.1.3 Concatenated Data

Due to the idiosyncratic nature of DNA amplification, only around a half of all the individuals sequenced for the cytochrome b locus were also sequenced for the D-loop. However, the 49 concatenated sequences (of 1059bp in length) may still prove informative, with a greater emphasis on nucleotide information content relative to geographical sampling effort. From this sub-sample of roach, the concatenated sequences yielded ten unique haplotypes from over 13 populations representing 11 rivers (Table 3.8 and Fig 3.4). As with the individual loci, the majority of individuals bore the haplotype

that was by far the most frequent within the general population, here named CON1 (N = 30, 61.22%). Unlike the individual locus haplotypes (in which the most common haplotype was found in all populations), CON1 was absent from one population: the Sussex Ouse. The second most commonplace haplotype – CON4 (N = 7, 14.29%) – was found in the southern rivers of the Thames, Kent Stour, Medway and Test, and in the Severn. The third most common haplotype - CON3 - is found in three individuals, two from the Kent Stour and a single roach from the River Medway. There are two individuals each that possess haplotypes CON8 and CON9, all four roach being found in the same river: the Sussex Ouse. The remaining haplotypes are all private: CON2 is found in the Great Ouse, CON5, CON6 and CON7 are found in the Suffolk Stour, Thames and Tees, respectively, whilst CON10 is also found in the Thames (River Ash).

Table 3.9 lists both haplotypic and nucleotide diversities according to catchment size. The Medway holds the greatest haplotypic and nucleotide diversities (h = 0.8330±0.222, π = 0.0012±0.0004). A similarly high level of diversity is observed in the Kent Stour, a system that neighbours the Medway (h = 0.800±0.164, π = 0.0010±0.0003). However, zero diversity is observed within the Yare and Trent. No correlation is observed between estimates of diversity and catchment size. Fig 3.4 suggests that, the River Tees aside, most observed diversity (i.e. numbers of unique haplotypes per drainage) occurs in the south of England, particularly so in the southeast catchments that drain into the English Channel.

Table 3.8: Haplotype counts and estimates of concatenated D-loop and cytochrome b diversity for each of 13 sampling sites within the UK

Pop	Haplotype*										Σ	Hap diversity		Nuc diversity	
	CON1	CON2	CON3	CON4	CON5	CON6	CON7	CON8	CON9	CON10		*h*	s.d.	π	s.d.
GON	2	1									3	0.6670	0.3140	0.0020	0.0010
KS	2		2	1							5	0.8000	0.1640	0.0010	0.0003
SSH	1										1	0.0000	0.0000	0.0000	0.0000
SSS	7				1						8	0.2500	0.1800	0.0005	0.0004
YR	3										3	0.0000	0.0000	0.0000	0.0000
TT	2					1					3	0.6670	0.3140	0.0007	0.0003
TA	6			1						1	8	0.4640	0.2000	0.0008	0.0004
TEST	2			2							4	0.6670	0.2040	0.0007	0.0002
TEES	1						1				2	1.0000	0.5000	0.0010	0.0005
SV	2			1							3	0.6670	0.3140	0.0007	0.0003
SO								2	2		4	0.6670	0.2040	0.0007	0.0002
SE	1										1	0.0000	0.0000	0.0000	0.0000
MP	1		1	2							4	0.8330	0.2220	0.0012	0.0004
Σ	30	1	3	7	1	1	1	2	2	1	49	**0.5140†**	**0.2012**	**0.0007**	**0.0003**

*Haplotype colours correspond to those of the pie chart segments in Fig 3.4.
†Mean values in bold type.

Table 3.9: Concatenated mtDNA diversity data by catchment (in decreasing order of size). N = number of samples; H = number of haplotypes; area is km^2.

Pop	Area	N	H	Hap diversity		Nuc diversity	
				h	s.d	π	s.d.
Thames	13513	11	4	0.4910	0.1750	0.0007	0.0003
Severn	11381	3	2	0.6670	0.3140	0.0007	0.0003
Trent	10329	1	1	0.0000	0.0000	0.0000	0.0000
Great Ouse	8443	3	2	0.6670	0.3140	0.0020	0.0010
Yare	3017	3	1	0.0000	0.0000	0.0000	0.0000
Suffolk Stour	2806	9	2	0.2220	0.1660	0.0005	0.0003
Kent Stour	2720	5	3	0.8000	0.1640	0.0010	0.0003
Tees	1792	2	2	1.0000	0.5000	0.0010	0.0005
Medway	1608	4	3	0.8330	0.2220	0.0012	0.0004
Test	1206	4	2	0.6670	0.2040	0.0007	0.0002
Sussex Ouse	574	4	2	0.6670	0.2040	0.0007	0.0002
μ	5217	3.8	2	**0.5467**	**0.2057**	**0.0008**	**0.0003**

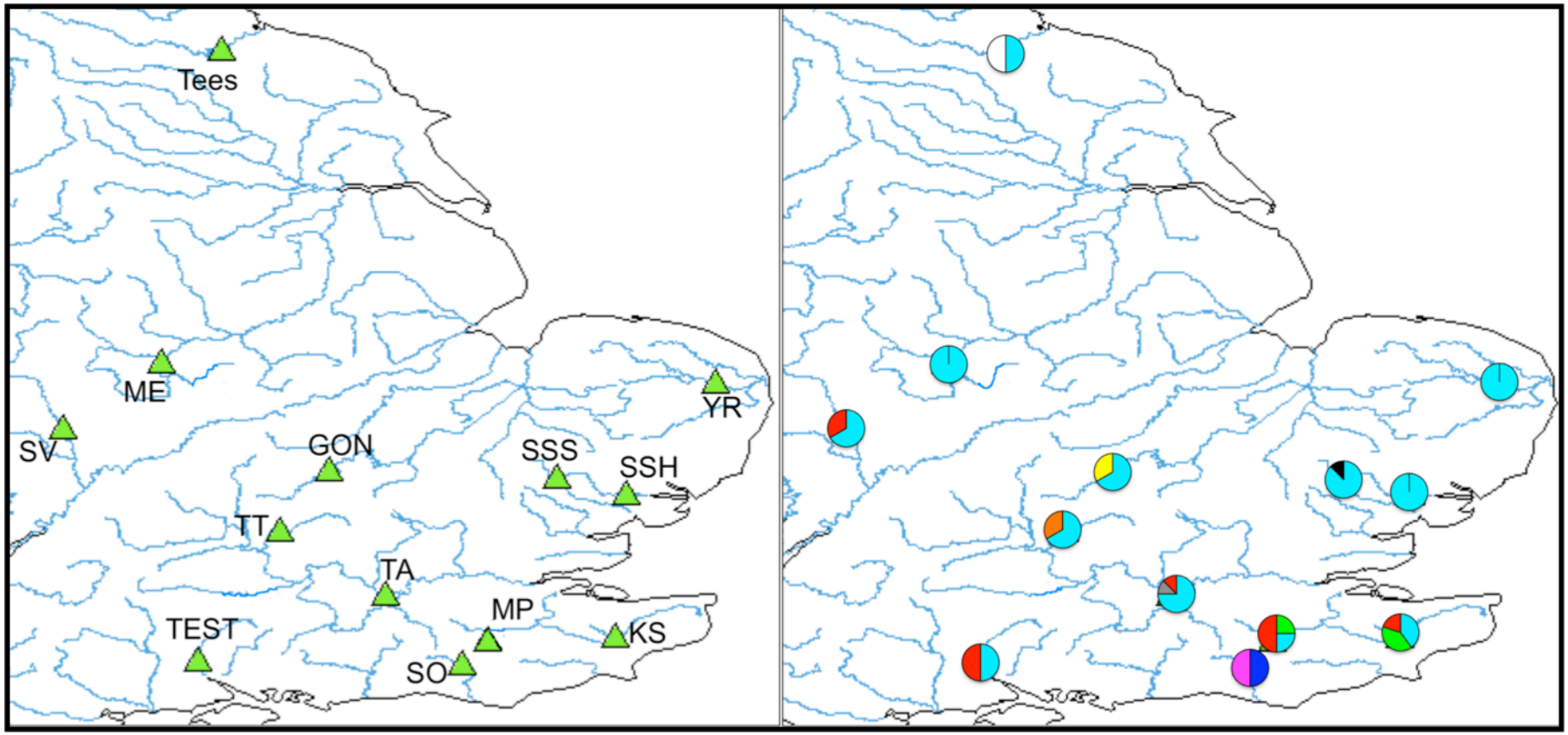


Figure 3.4: A frequency distribution map of concatenated mtDNA haplotypes. Right: Haplotype colours correspond to those in Table 3.8; Left: Sample site reference map.

3.3.2 Genetic Structuring

3.3.2.1 D-Loop

The analysis of molecular variance (AMOVA) of D-loop differentiation among individual populations and among rivers (Table 3.10) confirms that the majority of D-loop variation is to be found within populations (% variation = 87.82, Φ_{ST} = 0.123, *p*-value <0.001). Furthermore, significant differentiation is to be found within those rivers in which samples were taken from multiple sites (% variation = 6.99, Φ_{SC} = 0.074, *p*–value <0.001). However, the proportion of the variance in haplotype frequencies amongst rivers failed to reach significance, although 5% of all the variance in frequencies could be attributed to the influence of river systems upon levels of D-loop differentiation (% variation = 5.29, Φ_{CT} = 0.053, *p*-value = 0.097).

Table 3.10. Analysis of molecular variance (AMOVA) results for the D-loop dataset, where populations are grouped into rivers of origin at the highest level.

Source of variation	Degrees of freedom	Sum of squares	Variance	% of variation	Fixation indices	*p*-value
Among rivers	14	23.73	0.0132	5.29	Φ_{CT} = 0.0529	0.097
Among populations within rivers	12	6.80	0.0174	6.99	Φ_{SC} = 0.0738	<0.001
Within populations	1101	240.95	0.2189	87.82	Φ_{ST} = 0.1228	<0.001

Levels of genetic differentiation among all 26 *a priori* sample sites are given in Table 3.11. Above the diagonal partition are given estimates of F_{ST}; below diagonal are significance levels outputted from the G-statistic analysis. Of the 26 sites, only two showed consistently high levels of genetic differentiation with most other locations, although no one population was significantly divergent from all others. The River Severn sample is the most divergent (mean F_{ST} = 0.316±0.157, range of F_{ST} from 0.104 (Yare at Coston, not significant) – 0.773 (River Witham, highly significant, *p*-value << 0.001)). Of the 25 pairwise combinations involving the Severn population, 18 are highly significant after Bonferroni correction (72%), with another five significant populations if the number of pairwise comparisons is ignored. The exact same pattern is observed for the G-statistic.

The second most divergent population is the Sussex Ouse (mean F_{ST} = 0.229±0.032, range of F_{ST} from 0.154 (Great Ouse at Newport Pagnell, 0.05 <*p*> 0.00015) to 0.668 (River Witham, highly significant, *p*-value << 0.001)). The Sussex Ouse is significantly divergent from 13 populations in total (50%), rising to 21 if the Bonferroni correction is not applied. Again, the G-statistics mirror the F_{ST} significances, although only ten populations are considered statistically significant after correction.

Zero or negative F_{ST} values indicate a lack of spatial heterogeneity given by the *F*-statistics. However, little concordance is observed between genetic similarity and geographic proximity, either within rivers or among locally proximate sets of rivers. Differentiation within the Great Ouse is low (F_{ST} = 0.007 and not significant (n.s.); G-statistic *p*-value = 0.415), as well as within the Thames, with the exception for the River Kennet, which is significantly differentiated from the River Wey (F_{ST} = 0.298, *p*-value << 0.001; G-statistic *p*-value < 0.001), and also probably differentiated from the Rivers Thames and Ash (F_{ST} = 0.122 and 0.156, respectively, 0.05 < *p*-value > 0.00015 for F_{ST} and G-statistic), but not the Thame tributary at Shabbington (F_{ST} = 0.098, n.s.; G-statistic *p*-value = 0.147). Differentiation is also present but not significant between the Suffolk Stour samples (F_{ST} = 0.133, n.s.; G-statistic *p*-value = 0.094), within the Medway (F_{ST} = 0.044, n.s.; G-statistic *p*-value = 0.453), although uncorrected significance is found between the two Yare populations (F_{ST} = 0.555, 0.05 < *p*-value > 0.00015; G-statistic *p*-value = 0.013). Within the Trent, the two upper tributary populations of the Mease and Sence are undifferentiated (F_{ST} = -0.061, n.s.; G-statistic *p*-value = 0.643), although the Sence population is significantly differentiated from the Derwent sample before correction, while the Mease sample is not (F_{ST} = 0.236, < 0.05 *p*-value >0.00015; G-statistic *p*-value = 0.020; F_{ST} = 0.077, n.s.; G-statistic *p*-value = 0.127).

When within-river populations are pooled (Table 3.12), the Thames metapopulation is only statistically significantly divergent from the Kent Stour (F_{ST} = 0.108, *p*-value << 0.001; G-statistic *p*-value = << 0.001), the Severn (F_{ST} = 0.403, *p*-value << 0.001; G-statistic *p*-value = << 0.001), the Sussex Ouse (F_{ST} = 0.249, *p*-value << 0.001; G-statistic *p*-value = << 0.001) and the River Witham (F_{ST} = 0.123, *p*-value << 0.001; G-statistic *p*-value = 0.024). Tentative differentiation between the Thames and the Suffolk Stour may exist (F_{ST} = 0.056, 0.05 > *p*-value < 0.0005; G-statistic *p*-value = < 0.001).

Table 3.11: Pairwise F_{ST} and G-statistic significance levels between all sampled populations (below and above diagonal, respectively). Highlighted, underlined values (yellow) indicate statistical significance after correction for multiple comparisons (α= 0.05). Shaded values (grey) are those that are only significant before Bonferroni correction.

	YOB	YOC	GOE	GON	TH	TA	TT	TK	TW	SV	BA	SSH	SSS	YR	YC	HA	KS	W	TEES	DE	SO	TES	ME	SE	MP	MA
YOB		0.043	0.260	0.256	0.123	0.173	0.218	<0.001	0.143	<0.001	0.404	1.000	0.384	0.642	0.010	0.007	<0.001	0.339	0.147	0.705	<0.001	0.071	0.100	0.014	0.020	0.395
YOC	0.087		0.410	0.241	0.185	0.190	0.749	0.010	0.102	<0.001	0.104	0.020	0.034	0.032	0.313	0.069	<0.001	0.002	0.339	0.178	0.001	0.712	0.762	0.317	0.290	0.475
GOE	0.043	-0.020		0.415	0.546	0.212	0.625	0.002	0.390	<0.001	0.225	0.240	0.045	0.152	0.291	0.099	0.010	0.008	0.360	0.345	<0.001	0.755	0.807	0.475	0.273	0.516
GON	0.096	-0.005	0.007		0.084	0.184	0.685	<0.001	0.057	<0.001	0.114	0.210	0.014	0.045	0.298	0.023	<0.001	<0.001	0.049	0.394	0.023	0.703	0.784	0.337	0.490	0.813
TH	0.078	-0.015	-0.015	0.002		0.556	0.454	0.001	0.268	<0.001	0.361	0.146	0.006	0.106	0.373	0.228	<0.001	0.001	0.023	0.174	<0.001	0.865	1.000	0.558	0.324	0.662
TA	0.065	-0.002	-0.002	0.012	-0.012		0.289	0.004	0.050	<0.001	0.735	0.132	0.020	0.037	0.079	0.332	<0.001	0.005	0.043	0.257	<0.001	0.494	0.446	0.089	0.117	0.563
TT	0.073	-0.060	-0.062	-0.033	-0.049	-0.036		0.147	0.230	<0.001	0.254	0.085	0.128	0.126	0.247	0.333	0.037	0.030	0.449	0.272	0.020	1.000	1.000	0.434	0.486	0.806
TK	0.325	0.116	0.167	0.110	0.122	0.156	0.098		<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.743	0.033	<0.001	<0.001	<0.001	<0.001	<0.001	0.445	0.248	0.212	0.433	0.124
TW	0.003	0.045	-0.004	0.080	0.040	0.045	0.001	0.298		<0.001	0.286	0.278	0.018	0.611	0.046	0.011	<0.001	0.093	0.075	0.200	<0.001	0.230	0.569	0.070	0.046	0.189
SV	0.674	0.458	0.501	0.427	0.424	0.521	0.480	0.210	0.625		<0.001	<0.001	<0.001	<0.001	0.161	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.018	0.003	0.023	0.016	<0.001
BA	-0.003	0.034	0.002	0.059	0.029	-0.011	-0.005	0.264	-0.011	0.615		0.495	0.043	0.456	0.034	0.081	<0.001	0.065	0.086	0.458	<0.001	0.160	0.238	0.034	0.035	0.338
SSH	0.000	0.156	0.097	0.145	0.121	0.126	0.204	0.382	0.039	0.719	0.040		0.094	1.000	0.010	0.012	<0.001	1.000	0.201	0.456	<0.001	0.068	0.102	0.014	0.009	0.218
SSS	0.044	0.026	0.023	0.040	0.037	0.032	-0.007	0.209	0.064	0.546	0.046	0.133		0.036	0.018	0.002	<0.001	0.004	0.006	0.112	<0.001	0.114	0.137	0.023	0.020	0.320
YR	-0.020	0.125	0.060	0.132	0.098	0.102	0.121	0.379	-0.005	0.702	0.010	-0.016	0.113		0.013	0.002	<0.001	0.437	0.080	0.592	<0.001	0.098	0.145	0.019	0.014	0.288
YC	0.474	0.077	0.134	0.083	0.069	0.168	0.060	-0.025	0.354	0.104	0.329	0.633	0.219	0.555		0.505	0.029	0.002	0.176	0.015	0.010	0.591	0.331	0.643	1.000	0.185
HA	0.175	0.019	0.041	0.024	0.008	0.008	-0.006	0.049	0.138	0.348	0.094	0.231	0.089	0.215	0.011		<0.001	<0.001	0.011	0.019	<0.001	0.673	0.597	0.336	0.424	0.181
KS	0.219	0.092	0.099	0.057	0.091	0.122	0.074	0.136	0.202	0.395	0.182	0.260	0.148	0.253	0.113	0.077		<0.001	<0.001	0.001	<0.001	0.115	0.075	0.017	0.330	0.052
W	0.040	0.245	0.160	0.208	0.170	0.201	0.340	0.480	0.079	0.773	0.085	0.000	0.199	0.010	0.759	0.312	0.318		<0.001	0.301	<0.001	0.025	0.036	0.002	0.002	0.124
TEES	0.081	-0.015	0.002	0.003	0.008	0.016	-0.031	0.138	0.063	0.446	0.048	0.124	0.040	0.109	0.108	0.037	0.068	0.173		0.314	<0.001	0.543	0.634	0.225	0.376	0.392
DE	-0.030	0.065	0.028	0.079	0.064	0.047	0.044	0.298	-0.004	0.657	-0.012	0.013	0.062	-0.017	0.427	0.153	0.202	0.063	0.066		0.001	0.093	0.127	0.020	0.029	0.045
SO	0.459	0.231	0.292	0.154	0.246	0.265	0.228	0.213	0.443	0.538	0.387	0.553	0.300	0.540	0.267	0.208	0.242	0.660	0.225	0.424		0.022	0.010	0.003	0.014	0.030
TES	0.190	-0.058	-0.056	-0.023	-0.059	-0.004	-0.097	0.041	0.055	0.372	0.065	0.358	0.034	0.245	-0.074	-0.036	0.064	0.514	-0.019	0.152	0.237		1.000	1.000	0.754	0.467
ME	0.105	-0.051	-0.065	-0.012	-0.049	-0.015	-0.096	0.099	-0.007	0.450	0.013	0.239	0.017	0.136	0.025	-0.004	0.088	0.374	-0.014	0.077	0.267	-0.133		0.643	0.379	0.437
SE	0.271	0.007	0.026	0.032	0.001	0.073	-0.021	0.036	0.164	0.287	0.164	0.386	0.112	0.326	-0.104	-0.004	0.094	0.520	0.042	0.236	0.258	-0.117	-0.061		0.557	0.220
MP	0.359	0.008	0.047	-0.006	0.002	0.082	-0.017	-0.035	0.262	0.229	0.229	0.514	0.131	0.448	-0.143	-0.042	-0.030	0.663	0.015	0.311	0.193	-0.100	-0.027	-0.097		0.453
MA	0.061	-0.048	-0.044	-0.049	-0.063	-0.071	-0.086	0.136	0.024	0.564	-0.017	0.284	-0.017	0.146	0.159	0.021	0.088	0.455	-0.033	0.027	0.232	-0.027	-0.043	0.059	0.044	

Table 3.12: Pairwise F_{ST} and G-statistic significance levels between river systems (below and above diagonal, respectively). Highlighted, underlined values (yellow) indicate statistical significance after correction for multiple comparisons (α= 0.05). Shaded values (grey) are those that are only significant before Bonferroni correction.

	Thames	Trent	Great Ouse	Yorks Ouse	Yare	Suffolk Stour	Medway	Bristol Avon	Hamps Avon	Kent Stour	Severn	Sussex Ouse	Tees	Test	Witham
Thames		0.413	<0.001	0.03	0.66	<0.001	0.36	0.654	0.219	<0.001	<0.001	<0.001	<0.001	0.088	0.024
Trent	0.002		0.273	0.16	0.737	0.008	0.217	0.097	0.017	<0.001	<0.001	<0.001	0.028	0.675	0.012
Great Ouse	0.006	0.006		0.134	0.478	0.002	0.681	0.318	0.022	<0.001	<0.001	0.004	0.056	0.845	0.028
Yorks Ouse	0.001	0.001	0.014		0.452	0.131	0.138	0.439	0.003	<0.001	<0.001	<0.001	0.108	0.407	0.024
Yare	0.027	-0.013	0.028	-0.008		0.04	0.06	0.236	0.008	<0.001	<0.001	<0.001	0.052	0.282	0.053
Suffolk Stour	0.056	0.042	0.047	-0.001	0.022		0.012	0.052	<0.001	<0.001	<0.001	<0.001	<0.001	0.058	0.062
Medway	-0.031	-0.004	-0.034	0.036	0.063	0.102		0.091	0.293	0.101	<0.001	0.006	0.374	1	0.002
Bristol Avon	0.027	0.015	0.028	-0.01	-0.007	0.01	0.073		0.081	<0.001	<0.001	<0.001	0.086	0.16	0.065
Hamps Avon	0.013	0.061	0.033	0.097	0.12	0.154	-0.023	0.094		<0.001	<0.001	<0.001	0.011	0.674	<0.001
Kent Stour	0.108	0.143	0.076	0.162	0.19	0.207	0.015	0.182	0.077		<0.001	<0.001	<0.001	0.115	<0.001
Severn	0.403	0.503	0.441	0.573	0.599	0.63	0.387	0.615	0.348	0.395		<0.001	<0.001	0.018	<0.001
Sussex Ouse	0.249	0.334	0.212	0.345	0.421	0.412	0.195	0.387	0.208	0.242	0.538		<0.001	0.022	<0.001
Tees	0.016	0.027	0.001	0.024	0.054	0.068	-0.023	0.048	0.037	0.068	0.446	0.225		0.543	<0.001
Test	-0.052	-0.06	-0.036	0.018	0.017	0.097	-0.095	0.065	-0.036	0.064	0.372	0.237	-0.019		0.025
Witham	0.123	0.151	0.134	0.083	0.098	0.083	0.378	0.085	0.312	0.318	0.73	0.668	0.173	0.514	

Both the Severn and the Bristol Avon share the same ancestral connecting river – and share the same estuary – and yet the F_{ST} of 0.615 ($p << 0.001$) shared between these populations is the fourth highest value of divergence in all the pairwise comparisons. However, the geographically proximate Hampshire Test and Avon are indistinguishable from each other by their haplotype frequencies (F_{ST} = -0.036, n.s.; G-statistic p-value = 0.674). The Trent and the Yorkshire Ouse share the Humber estuary, by which they currently – and historically – drain into the North Sea. These two rivers are also indistinguishable by traditional F-statistics or by the distribution of haplotype frequencies (F_{ST} = 0.001, n.s.; G-statistic p-value = 0.160). Similarly, the mouths of the Medway and Kent Stour are proximal to one another, and they also show minimal divergence (F_{ST} = 0.015, n.s.; G-statistic p-value = 0.101). The Witham maintains mostly non-zero F_{ST} scores by the virtue of being fixed for a single haplotype (mean F_{ST} = 0.199±0.17), even with the Great Ouse with which it shares the Wash as a point of discharge (F_{ST} = 0.134, p-value << 0.001; G-statistic p-value = 0.028).

3.3.2.2 Cytochrome B

The more modest cytochrome b dataset reveals a different pattern of hierarchical structuring according to the AMOVA analysis (Table 3.13), although all inferences of hierarchical structure are unsupported statistically. Whilst most variation is held within populations (92.29%, Φ_{ST} = 0.077, p-value = 0.098), no variation was apportioned to among population differences within rivers, and around 7.8% was apportioned among rivers themselves, albeit with little statistical support (Φ_{CT} = 0.097, p-value = 0.168).

Table 3.13: Analysis of molecular variance (AMOVA) results for the cytochrome b dataset, where populations are grouped into rivers of origin at the highest level.

Source of variation	Degrees of freedom	Sum of squares	Variance	% of variation	Fixation indices	*p*-value
Among rivers	10	1.465	0.009	9.66	Φ_{CT} = 0.097	0.168
Among populations within rivers	5	0.360	-0.002	-1.96	Φ_{SC}= -0.022	0.410
Within rivers	81	6.876	0.085	92.29	Φ_{ST} = 0.077	0.098

Table 3.14: Pairwise F_{ST} values (below diagonal) and G-statistic significance (above diagonal) for the cytochrome b dataset.

	GON	SE	ME	SSH	SSS	YR	TEES	KS	MP	MA	TA	TT	TK	TEST	SO	SV
GON		N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.
SE	-0.0588		N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.
ME	0.2500	-0.0262		N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.
SSH	0.0000	-0.0843	0.1892		N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.
SSS	-0.0645	0.0005	0.1951	-0.0891		N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.
YR	0.0000	-0.1842	0.0000	0.0000	-0.1871		N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.
TEES	0.5385	0.3860	0.0455	0.4737	0.4161	0.2500		N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.
KS	0.0000	-0.0588	0.2500	0.0000	-0.0645	0.0000	0.5385		N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.
MP	0.0000	-0.0843	0.1892	0.0000	-0.0891	0.0000	0.4737	0.0000		N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.
MA	0.0000	-0.3235	-2.0000	0.0000	-0.3250	0.0000	0.0000	0.0000	0.0000		N.S.	N.S.	N.S.	N.S.	N.S.	N.S.
TA	-0.0588	0.0000	0.1725	-0.0843	0.0005	-0.1842	0.3860	-0.0588	-0.0843	-0.3235		N.S.	N.S.	N.S.	N.S.	N.S.
TT	0.0000	0.0191	0.4667	0.0000	0.0083	0.0000	0.7303	0.0000	0.0000	0.0000	0.1907		N.S.	N.S.	N.S.	N.S.
TK	-0.0588	-0.0714	-0.0787	-0.0843	0.0037	-0.1842	0.2230	-0.0588	-0.0843	-0.3235	0.0000	0.0191		N.S.	N.S.	N.S.
TEST	0.0000	-0.1215	0.1111	0.0000	-0.1253	0.0000	0.3846	0.0000	0.0000	0.0000	-0.1215	0.0000	-0.1215		N.S.	N.S.
SO	0.2941	0.1139	-0.3505	0.2500	0.2923	0.1177	0.2019	0.2941	0.2500	-0.0170	0.2734	0.4619	0.0625	0.1946		N.S.
SV	0.0000	-0.1842	0.0000	0.0000	-0.1871	0.0000	0.2500	0.0000	0.0000	0.0000	-0.1842	0.0000	-0.1842	0.0000	0.1177	

Population divergence as recorded by cytochrome b haplotype frequencies are given in Table 3.14. Not a single comparison was significant, indicating negligible differentiation among populations. In some cases, however, F_{ST} was high (e.g. most River Tees comparisons), although one must conclude that low sample size in conjunction with low sequence variation has contributed to an erosion of statistical significance across all cytochrome b structuring analyses.

3.3.2.3 Concatenated Data

Because of the limited number of within river sites (the degrees of freedom for the portion of total variance in haplotype frequencies assigned to the differences among populations within rivers = 2) and limited sampling, statistical power is compromised, particularly when assessing for differences among populations within rivers (Table 3.15). However, this analysis – like the others - suggests that most variation in haplotype frequency is to be found within sites (88.38%, Φ_{ST} = 0.310, *p*-value = 0.074). The remaining variation is largely apportioned among rivers (11.62%, Φ_{CT} = 0.311, *p*-value = 0.101), after accounting for the negative artifact association with Φ_{SC}.

Table 3.15: Analysis of molecular variance (AMOVA) results for the concatenated dataset, where populations are grouped into rivers of origin.

Source of variation	Degrees of freedom	Sum of squares	Variance	% of variation	Fixation indices	*p*-value
Among rivers	10	4.565	0.093	31.10	Φ_{CT} = 0.311	0.101
Among sites within rivers	2	0.177	-0.060	-19.48	Φ_{SC} = -0.283	1.000
Within sites	36	13.958	0.274	88.38	Φ_{ST} = 0.310	0.074

Table 3.16: Pairwise F_{ST} (below diagonal) and G-statistic significance (above diagonal) for the concatenated data. Shaded, grey values are significant before Bonferroni correction. No other values approach significance.

	GON	KS	SSH	SSS	YR	TT	TA	TEST	TEES	SV	SO	SE	MP
GON		n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
KS	-0.0139		n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
SSH	-1.0000	-0.3330		n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
SSS	0.0188	0.2469	-1.0000		n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.0244	n.s.	0.0362
YR	0.0000	0.2053	0.0000	-0.1748		n.s.	n.s.	n.s.	n.s.	n.s.	0.0019	n.s.	n.s.
TT	-0.2000	-0.0139	-1.0000	0.0188	0.0000		n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
TA	-0.0900	0.0880	-0.8571	-0.0390	-0.1103	-0.0900		n.s.	n.s.	n.s.	0.0068	n.s.	n.s.
TEST	0.0000	-0.0526	-0.3333	0.2615	0.2500	0.0000	0.0204		n.s.	n.s.	n.s.	n.s.	n.s.
TEES	-0.2000	-0.0825	-1.0000	0.2178	0.2500	-0.2000	0.0109	-0.0435		n.s.	n.s.	n.s.	n.s.
SV	-0.2000	-0.1135	-1.0000	0.0188	0.0000	-0.2000	-0.1967	-0.3333	-0.2000		n.s.	n.s.	n.s.
SO	0.3333	0.2614	0.3333	0.6033	0.6129	0.3333	0.4607	0.3333	0.2258	0.3333		n.s.	n.s.
SE	-1.0000	-0.3333	0.0000	-1.0000	0.0000	-1.0000	-0.8571	-0.3333	-1.0000	-1.0000	0.3333		n.s.
MP	0.0891	-0.1656	-0.1111	0.3969	0.3684	0.0891	0.1804	-0.2000	-0.0182	-0.1358	0.2500	-0.1111	

Pairwise estimates of population differentiation for the limited concatenated dataset failed to uncover any significant structuring (Table 3.16). Not a single pairwise comparison was significant after Bonferroni correction, although four were prior to correction. Three of these four borderline comparisons were between the Sussex Ouse and the upper Suffolk Stour (F_{ST} = 0.603), the Yare at Rockland (F_{ST} = 0.613) and the River Ash (F_{ST} = 0.461), with the upper Suffolk Stour also divergent from the Medway at Penshurst (F_{ST} = 0.397).

3.3.3 Genealogy and Demographic History of D-Loop Lineages

For all included D-loop sequences, the AIC criterion within jModelTest found that a TIM3+I+G model of nucleotide substitution (a transition model) best suited the data (-Ln = 1172.27, proportion of invariant sites = 0.882, gamma α shape = 0.692). All D-loop phylogenetic analyses incorporated these parameters.

Figs 3.5a and 3.5b display 50% majority-rule consensus trees calculated by the MP and ML algorithms, respectively. Including European roach sequences (N = 148; see Appendix A.2 for details), a total of 40 unique haplotypes were observed for the data, ten of which are clustered monophyletically within *R. r. caspicus* (Haplotypes D19 – D30). The phylogenetic trees broadly agree with one another with regards to their topologies, with strong support for reciprocal monophyly of the Caspian roach and common roach found using the MP algorithm (86% bootstrap support). The same relationship is found in the ML tree, although the support for the distinction between these two clades is much lower (44% bootstrap support, not shown). Both algorithms agree with the basal placing of haplotype D37 within the common roach clade (MP: 90% bootstrap support; ML: 86% bootstrap support). Given the shallow nature of the phylogeny, in addition to restricted sampling, strong bootstrap support was not expected or found for any other within-clade relationship.

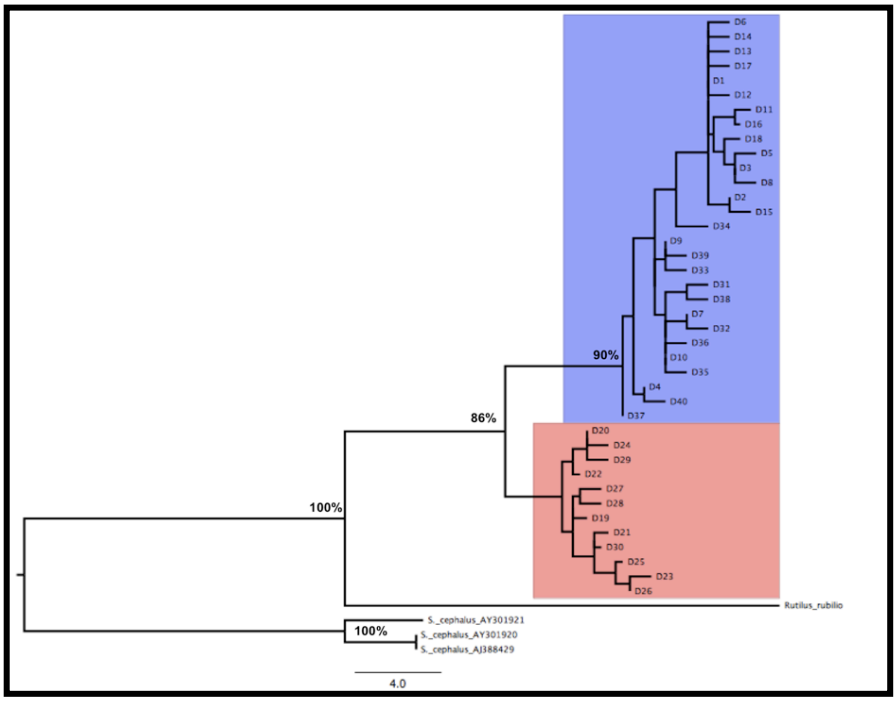


Figure 3.5a: 50% majority-rule consensus tree of phylogenetic relationships within roach based upon a maximum parsimony analysis of roach D-loop sequences. Bootstrap support (%) is only shown for those nodes that are supported in over 50% of the 1000 bootstrap replicates. Blue clade: *Rutilus rutilus*. Red clade: *Rutilus rutilus caspicus*. Scale: number of inferred nucleotide changes.

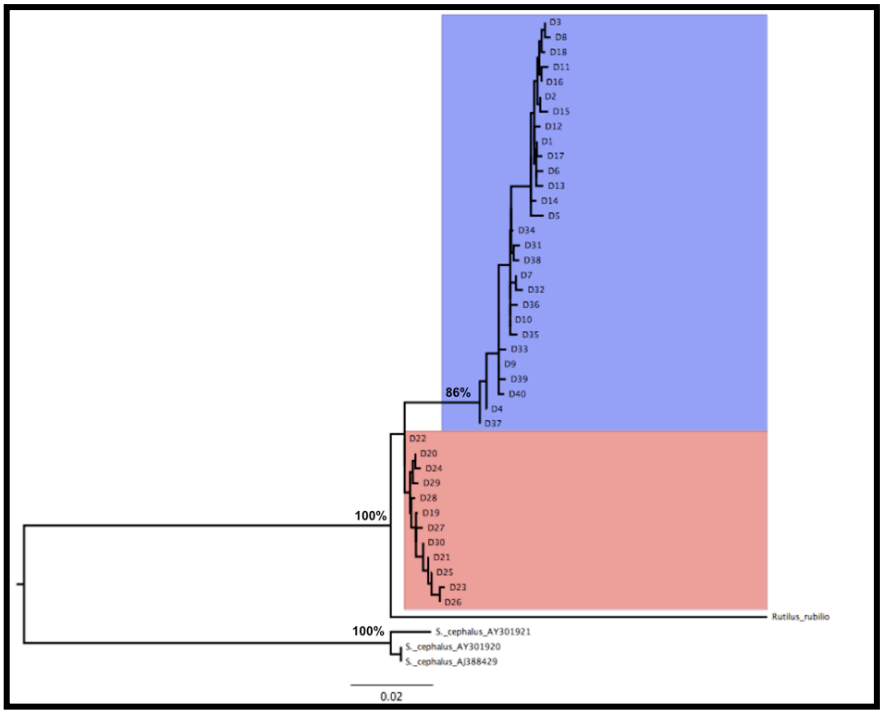


Figure 3.5b: 50% majority-rule consensus tree of phylogenetic relationships within roach based upon a maximum likelihood analysis of roach D-loop sequences. Bootstrap support (%) is only shown for those nodes that are supported in over 50% of the 1000 bootstrap replicates. Blue clade: *Rutilus rutilus*. Red clade: *Rutilus rutilus caspicus*. Scale: proportion of inferred nucleotide changes.

The phylogeny of the D-loop sequences found only within the UK is very shallow. Resolving the statistical parsimony network of UK haplotypes by standard coalescent theory-based reasoning produces a simple, star-like phylogeny whereby most unique haplotypes are seemingly derived from haplotype D1 by just one mutational step (Fig 3.6). Within this scheme, the greatest mutational distance between two haplotypes is five steps: between D4 or D9 and any of the D3-derivatives.

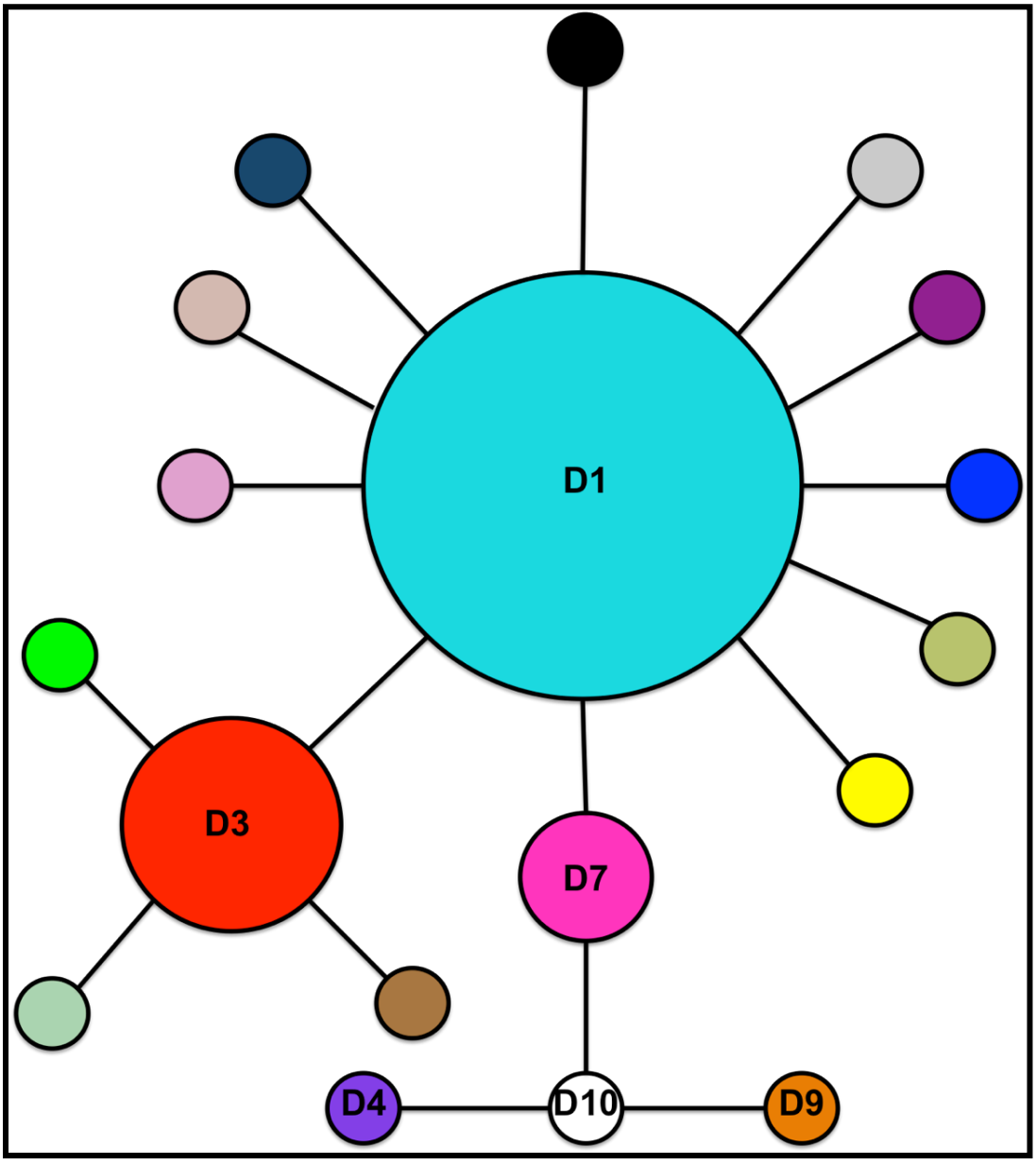


Figure 3.6: Network of D-loop haplotypes found within the UK, determined by statistical parsimony. All loops were resolved according to the criteria of Crandall & Templeton (1993). The three most common haplotypes are shown by circles of decreasing area, albeit not to scale, for ease of depiction. Colours correspond to those depicted for haplotypes D1-D18 in Table 3.3 and Fig 3.2.

A similar haplotypic network was derived using the MJN algorithm of Bandelt et al. (1999). As many loops as possible were eradicated by repeating the MJ-algorithm using differently weighted, variable nucleotide sites (hypervariable sites were weighted lower) (Fig 3.7). Adopting the coalescent criteria outlined above for the TCS analysis, the edges describing the 18 UK haplotypes of Fig 3.7 collapse further and resemble those of the network observed in Fig 3.6. Fig 3.8 shows a haplotype network constructed in Haplotype Viewer, using the maximum parsimony analysis of the data as primary information in determining genealogical relationships. For this latter analysis, polarity of nucleotide evolution was established by incorporating extra D-loop sequences from the common roach and from the Caspian roach.

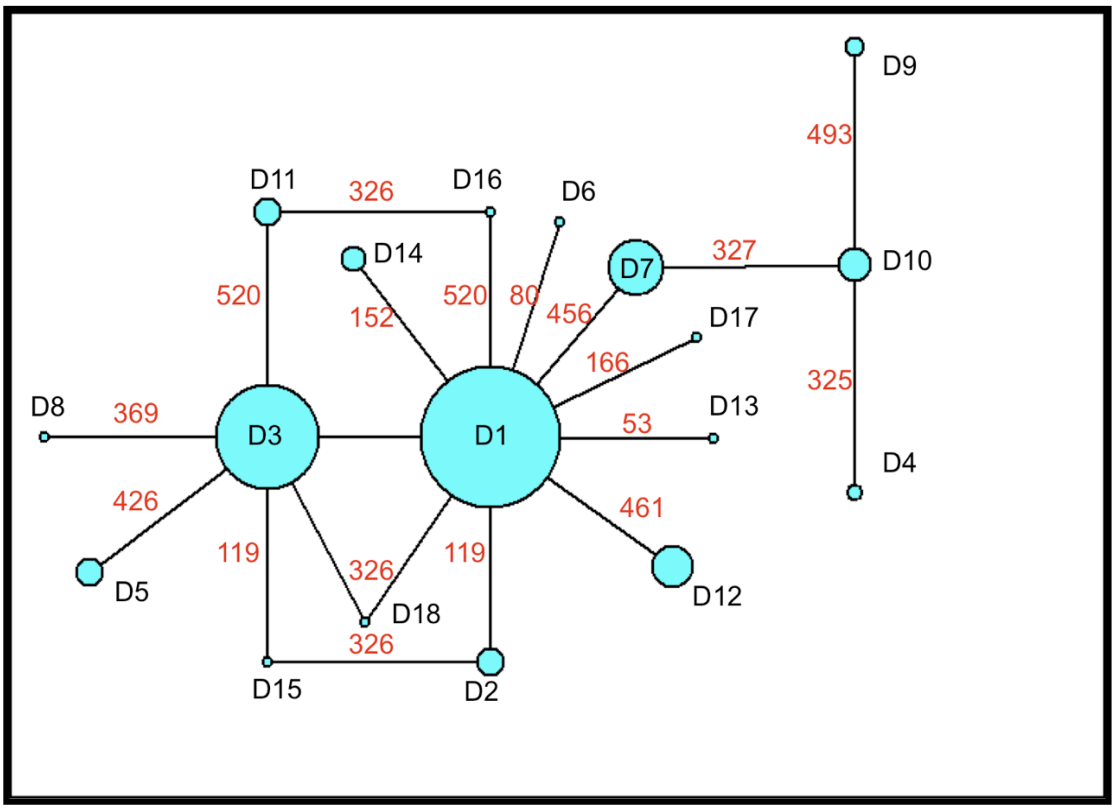


Figure 3.7: Network constructed using the MJN algorithm showing resolved (edges) and unresolved relationships (loops, or reticulations) among haplotypes. The red numbers indicate the nucleotide position at which mutations are inferred to have occurred.

Concerning *R. rutilus sensu stricto*, the MP tree-transformed network strongly suggests that haplotype D1 is *not* the ancestral type, instead inferring the ancestral type within the UK to be haplotype D4. The chain of mutation among basal haplotypes is generally consistent across methodologies: D10 is ancestral to D7, which itself begat D1. The methods deviate in ascribing some of the minor haplotypes to either D1 or D3, with the MP-method resulting in no reticulation of mutations, inferring that haplotypes D2 and D16

are descendants of D3 via D15 and D11, respectively, in addition to D18. The utilization of European sequences indicates that D4 is derived from D37, a haplotype found within the Danube and Rhine drainages in two individual roach.

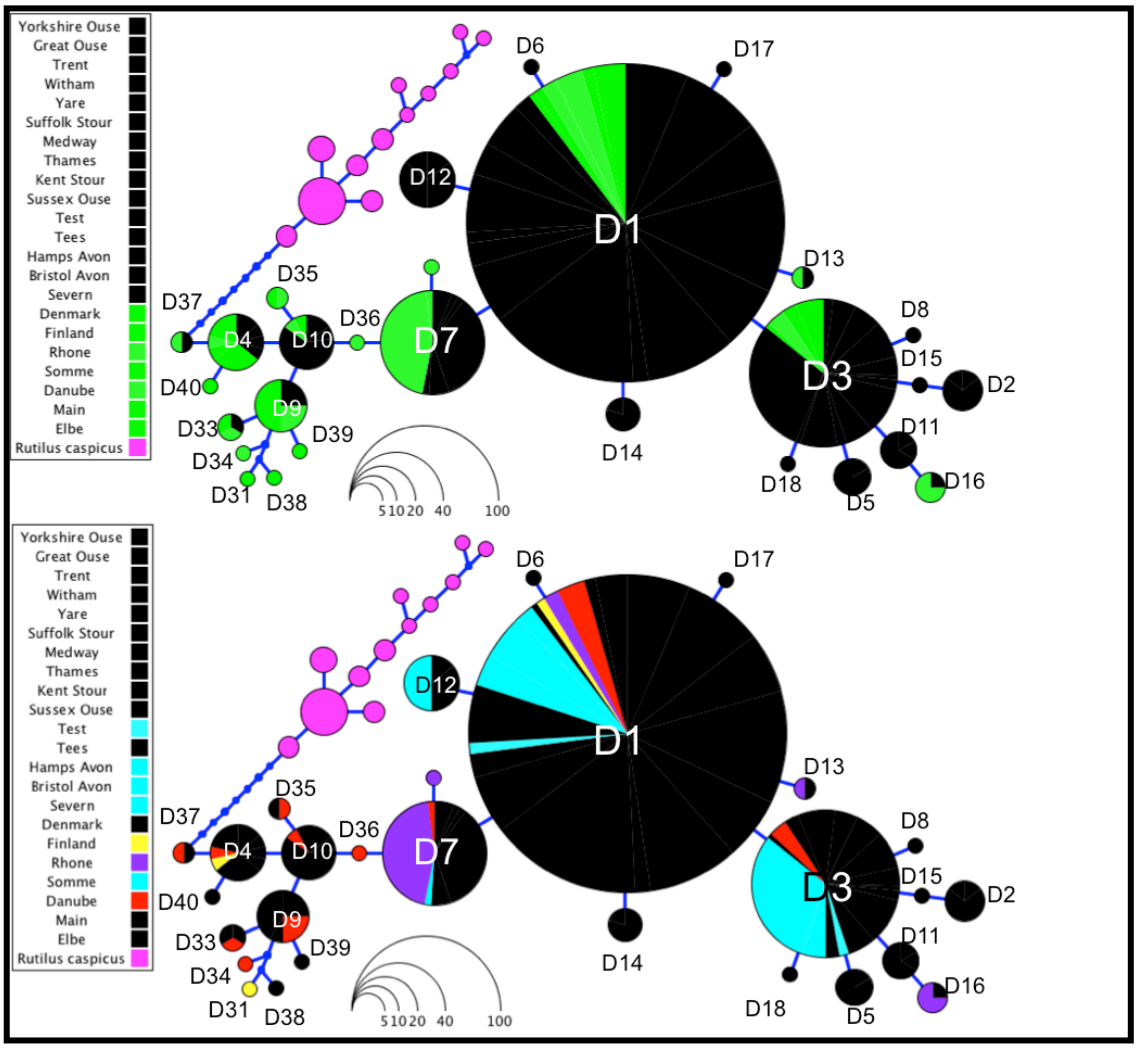


Figure 3.8: MP tree-transformed haplotype networks. Haplotypes are each represented by circles whose size is proportional to the number of individual bearers of that haplotype found within the sample of 712 roach, inclusive of European fish. Top: coloured segments represent the proportion of each haplotype found within either the UK (black) or on the European mainland (green). Bottom: colours represent the haplotypes found in rivers that discharge into particular marine territories: North Sea (black), English Channel (light blue), the Mediterranean (purple), the Baltic (yellow) and the Black Seas (red). For both figures *R. r. caspicus* haplotypes are shown in magenta.

Fig 3.8 displays the MP tree-transformation network in two forms: the upper panel shows UK versus European distributions of haplotypes; the lower panel apportions haplotypes to the discharge sea of their river of origin. There are three important observations: firstly, the UK possesses a number of ancestral haplotypes; second, much of the tip-diversification is associated with haplotypes D3 and D1, such that the majority of descendent haplotypes are mostly found in UK waters; thirdly, much of the D7 haplotype-

bearing roach, from which the vast majority of UK roach bear its descendent lineage, are located in the River Rhône, in the south of France, along with a few individuals of haplotypes D13 and D16. As expected, all of the most ancestral haplotypes are also found, sometimes exclusively, in the drainages of central Europe and the Danube in particular.

The phylogenetic reconstruction methods differ in their pictographic representations of the evolutionary process due to the presence of irreducible network loops that are impossible to eradicate completely using the weighting or MP option in the MJN analysis, yet are usually fully resolved using traditional phylogenetic inference. Furthermore, the MP network extends the maximum mutational distance to seven mutational steps (spanning D4/D9 to D16/D2) within the UK D-loop diversity. Additionally, The haplotype D36 - found in a single individual roach from the Danube - is placed in between haplotypes D10 and D7 in the tree-transformation network, whilst no intermittent placing of a similar haplotype is inferred from an analysis of UK haplotypes alone. This is due to an apparent reticulation that may also be resolved by allowing D36 to be a tip descendent of D10.

Table 3.17 lists the uncorrected p-distances obtained for pairwise estimates of nucleotide divergence between each of the UK haplotypes. The greatest distances are observed between haplotypes D11, D8 and D5 with the ancestral haplotype of UK roach, D4 (p-distance = 0.00789) and with haplotype D9. This finding casts doubt on whether haplotypes D16 or D2 are derived from haplotypes D15 and D11 (as inferred from an MP analysis, see Fig 3.8), respectively, as the distance between D16 and D4/D9 is 0.00691, as is the distance between D2 and D4/D9. Mean pairwise p-distance across all 18 UK haplotypes is 0.0047± 0.0025.

No overt association between haplotype, or genealogical grouping of haplotypes, was observed in the UK, or in the limited European sample, with the exception of the Caspian roach, which again confirmed their reciprocal monophyly (Figs 3.5.a and 3.5.b) as previously speculated by Ketmaier et al. (2008), and D7, which is prevalent in the Rhône. However, significant numbers of haplotypes are found *in situ* within the UK that are derived from haplotypes that are most frequently represented within the UK subset of all considered catchments (Fig 3.8).

Table 3.17: Pairwise p-distances among the 18 D-loop haplotypes found in the UK.

	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16	D17	D18
D1																		
D2	0.00158																	
D3	0.00158	0.00315																
D4	0.00473	0.00631	0.00631															
D5	0.00315	0.00473	0.00158	0.00789														
D6	0.00158	0.00315	0.00315	0.00631	0.00473													
D7	0.00158	0.00315	0.00315	0.00315	0.00473	0.00315												
D8	0.00315	0.00473	0.00158	0.00789	0.00315	0.00473	0.00473											
D9	0.00473	0.00631	0.00631	0.00315	0.00789	0.00631	0.00315	0.00789										
D10	0.00315	0.00473	0.00473	0.00158	0.00631	0.00473	0.00158	0.00631	0.00158									
D11	0.00315	0.00473	0.00158	0.00789	0.00315	0.00473	0.00473	0.00315	0.00789	0.00631								
D12	0.00158	0.00315	0.00315	0.00631	0.00473	0.00315	0.00315	0.00473	0.00631	0.00473	0.00473							
D13	0.00158	0.00315	0.00315	0.00631	0.00473	0.00315	0.00315	0.00473	0.00631	0.00473	0.00473	0.00315						
D14	0.00158	0.00315	0.00315	0.00631	0.00473	0.00315	0.00315	0.00473	0.00631	0.00473	0.00473	0.00315	0.00315					
D15	0.00315	0.00158	0.00158	0.00789	0.00315	0.00473	0.00473	0.00315	0.00789	0.00631	0.00315	0.00473	0.00473	0.00473				
D16	0.00158	0.00315	0.00315	0.00631	0.00473	0.00315	0.00315	0.00473	0.00631	0.00473	0.00158	0.00315	0.00315	0.00315	0.00473			
D17	0.00158	0.00315	0.00315	0.00631	0.00473	0.00315	0.00315	0.00473	0.00631	0.00473	0.00473	0.00315	0.00315	0.00315	0.00473	0.00315		
D18	0.00158	0.00315	0.00158	0.00631	0.00315	0.00315	0.00315	0.00315	0.00631	0.00473	0.00315	0.00315	0.00315	0.00315	0.00315	0.00315	0.00315	

As previously alluded, haplotype D3 is only found in significant numbers in a mostly southern and western distribution (Fig 3.2), as are most of the haplotypes inferred to be derived from it, with the exception of D5 which is found in the north and east (Rivers Tees and Yorkshire Ouse) and one individual each of D2 and D16, again in the Tees and Yorkshire Ouse, respectively. D3 and D16 are also found in central Europe and, for three individuals of D16 in particular, in the Rhône of southern France. The River Rhône is a repository of a large number of the D7 individuals (N = 22 (45% of all D7 roach)), where the D7 variant makes up 65% of all Rhône haplotypes (see Appendix A.2). The more ancient haplotypes D4, D7, D9 and D10 are mostly present only in the south of England, with the exception of three D7 roach in the Tees and, intriguingly, the only two representatives of the most ancient UK lineage haplotype -D4 - in the Yorkshire Ouse. The potential significance of these observations is discussed below.

Fig 3.9 shows the 95% HPD limits for the likelihood calculated for each model used to infer the phylogenetic relationships among ingroup and outgroups haplotypes. The mean log likelihood values for each model are as follows: constant: LnL = -1955.10 (95% HPD: -1968.25 to -1943.12); Yule: LnL = -1951.47 (95% HPD: -1963.84 to -1939.69); generalized skyline: LnL = - 1953.12 (95% HPD: -1965.66 to - -1940.86); and amalgamated Yule-skyline (Y-S): LnL = - 1952.79 (95% HPD: -1965.28 to -1940.91). Repeating the analyses for each model but sampling from only the prior confirmed that the prior itself was not driving the results. Bayes factors were calculated from the marginal likelihoods for each of the postulated demographic and speciation models. Log10 Bayes factors indicate the strength of support of one model over another when they are directly compared. For all comparisons, the log10 Bayes factors support the Yule model over the

two coalescent models. In accordance with Pfenniger et al.[18] the improvement in model fitting of the Yule model over a model of constant population size is strong (log10 BF = 1.658), however the strength of support of the Yule model over the skyline model is significant (log10 BF = 0.742) and borderline significant over the S-Y model (log10 BF = 0.512). Both the skyline model and the S-Y model are better suited to the data than a model of constant population size (log10 BF = 0.916 and 1.146, respectively). Of all the models, the amalgamated Y-S model provided most confidence in root placement (Y-S: mean = 26.20 Mya (95% HPD: 23.31 - 30.30 Mya); constant: mean = 27.67 Mya (95% HPD: 23.00 – 42.07 Mya); Yule: mean = 28.98 (95% HPD: 23.39 – 37.50 Mya); generalized skyline: mean = 26.96 (95% HPD: 23.36 – 32.18)) and for most of the other highlighted clades.

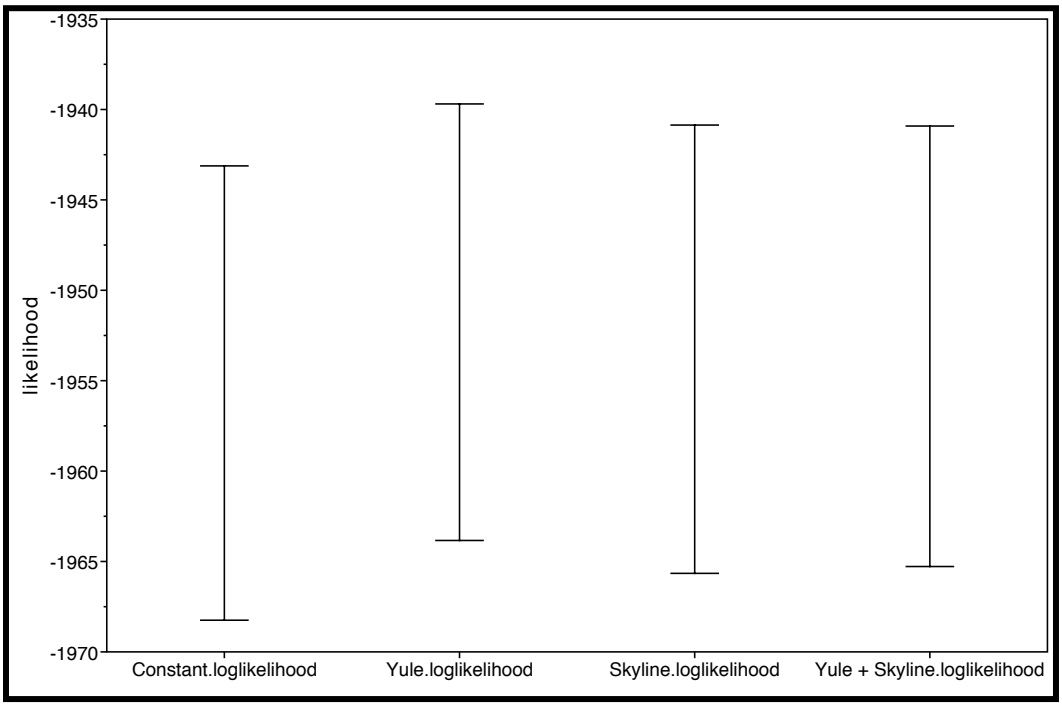


Figure 3.9: 95% HPD bars of the log likelihood distributions of one speciation (Yule), two coalescent (constant and skyline) and one hybrid (Yule and skyline) model underpinning phylogenetic within the BEAST software.

Given the preponderance of younger coalescences to older ones in the tree, a model that incorporates a coalescent model probably reflects a more accurate interpretation of the data than a strict speciation model of evolution that does not take into account past

[18] What constitutes significant BF support is not universal in the literature, depending on strength of data in information content and upon sampling effort, or to increased statistical rigidity due to an application to clinical study (e.g. only BF > 5 is considered significant evidence of improved model fitting in a study of the HIV virus (Hughes et al. 2009; Jeffreys 1967)). Pfenniger et al. (2010) working within crustaceans suggest that log10 Bayes factor >0.48 signifies substantial support, >1.00 strong support, and >1.48 very strong support.

demography and places disproportionately more statistical emphasis on older nodes. Mean mutation rate was, therefore, extracted from the skyline model. Tamura's test of mutation rate constancy among and within lineages (1992) as applied in MEGA 5.05 validated the utilization of a relaxed clock in the BEAST analyses. The supposition of a strict clock describing mutational events within the dataset was roundly rejected (-LnL clock: 1572.04; -LnL without clock: 1484.21; *p*-value << 0.001). The coefficient of variation among lineages in the BEAST analysis was 1.533 (95% HPD: 0.989 – 2.177), corroborating the log likelihood test. The reliability of the data is further supported by the lack of autocorrelation of rates within the tree (the covariance of parent and daughter lineages traverse zero: mean covariance = -0.0429 (95% HPD: -0.188 to 0.135)). The inferred mean nucleotide substitution rate averaged across the entire phylogeny (including the outgroup sequences) was 0.0073 (95% HPD: 0.0052 – 0.0096) nucleotide substitutions per site per million years, or 1.46% sequence divergence per million years.

The inferred substitution rate for the node at which the common roach and Caspian roach diverged was 0.0073 (95% HPD: 0.0001 – 0.0236) (1.46% sequence divergence per million years). The substitution rate inferred for the basal node in the common roach clade (when D37 diverged) is 0.035 (95% HPD: 0.0013 – 0.1056), or 7% sequence divergence per million years. The mean substitution rate calculated from all nodes within the common roach clade is 0.0291 substitutions per lineage per million years, or 5.82% sequence divergence per million years. Applying the relaxed clock to dating significant divergences within the rudimentary D-loop phylogeny, the common roach and Caspian roach became distinct clades some 927900 years ago (95% HPD: 0.167 – 2.110 Mya). The diversification of the common roach clade, based on surviving lineages, began approximately 0.292 Mya (95% HPD: 0.072 – 0.743 Mya). The split of all *Rutilus rutilus* spp with *Rutilus rubilio* was dated at 1.92 Mya (95% HPD: 0.318 – 3.545 Mya).

A number of tests were carried out to test the hypothesis of demographic stability. The results of these tests, and the values of the statistics upon which all or some are derived (e.g. S, θ and τ), can be seen in Table 3.18. The null hypothesis of population expansion could not be rejected for the entire *R. rutilus* clade ("All") (mean mismatch value = 1.816±3.50, sum of squared differences (SDD) = 0.006, *p*-value = 0.07), all pooled UK roach (mean mismatch value = 0.639±0.790, SSD = 0.001, *p*-value = 0.343), and clades D7 (mean mismatch value = 0.579±0.666, SSD = 0.005, *p*-value = 0.078) and D1 (mean mismatch value = 0.463±0.614, SSD = 0.002, *p*-value = 0.161). However, for all common roach and for the tip-clade, D3, population expansion was not supported (Roach s.s: mean

mismatch value = 0.899±1.110, SSD = 0.001, *p*-value = 0.046; D3: mean mismatch value = 0.580±0.814, SSD = 0.22, *p*-value < 0.001). However, for all groups, Harpending's r was not significantly different from expectations of a smooth, demographic expansion (see Table 3.18 and Fig 3.10). The neutrality analysis results are in general agreement with the mismatch analysis: For the all-inclusive dataset and for *R. rutilus sensu stricto*, population stasis was rejected (All: Fs = -14.032, *p*-value <<0.001, Tajima's D = -1.534, *p*-value = 0.030; Roach s.s: Fs = -26.752, *p*-value <<0.001, Tajima's D = -1.434, *p*-value = 0.024). Estimates of Fs were also found to be significant for the remaining groupings (UK roach: Fs = -15.39, *p*-value = < 0.001; D7: Fs = -12.902, *p*-value = 0.001, D1: Fs = -12.131, *p*-value = 0.002; D3: Fs = -5.466, *p*-value = 0.026), however, Tajima's D was borderline significant for clade D1 (D = -1.406, *p*-value = 0.045), but non-significant for clades D7 and D3 (D = -1.426, *p*-value = 0.138 and D = -0.777, *p*-value = 0.258, respectively). A more powerful statistic (R_2) was implemented to test whether the null hypothesis of population stasis, rather than expansion, holds true. Population stasis was rejected for all included roaches, common and Caspian ("All": R_2 = 0.027, *p*-value = 0.042) and for Roach *s.s*. (R_2 = 0.022, *p*-value = 0.05). For the clades D7 through D3, population stability could not be rejected: R_2 = 0.025 (*p*-value = 0.185) and 0.060 (*p*-value = 0.258) for D7 and D3 respectively, but D1 approached significance (R_2 = 0.024, *p*-value = 0.055). Taking the UK roach as a single entity, population expansion was not supported (R_2 = 0.023, *p*-value = 0.168; *r* = 0.101, *p*-value = 0.533).

By incorporating the mutation rates estimated from the BEAST analysis into Slatkin & Hudson's equation (1991), the time since expansion for each group listed in Table 3.18 was calculated. The time since expansion for each group, based on a maximum divergence rate of 7% per million years, a mean rate of 5.82% per million years within the common roach clade and a lowest inferred rate – and overall mean - of 1.46% per million years (nucleotide substitution rate of 0.035, 0.0291 and 0.0073 per site per lineage per million years) are as follows: All: 18809/22595/90281 years ago; Roach s.s: 20430/24541/97516 years ago; UK roach: 14512/17432/69654 years ago; D7: 13027/15649/62527 years ago; D1: 11384/13676/54644 years ago; and D3: 6592/7919/31641 years ago. The value of τ for Roach s.s. (0.908) is greater than that for all roaches (0.836) resulting in an extended time to coalescence for a younger node, probably due to the fact that two evolutionary distinct lineages are combined. Even so, the time since expansion for the ancestor of the common roach is dated at around 20-90 thousand years ago. The expansion of the D3 clade is synchronous with the isolation of the British Isles from the European mainland via

the formation of the English Channel if one adopts the mean rate of nucleotide substitution within the common roach clade, but not if the overall mean rate is applied.

Table 3.18: Statistical analyses of demographic history performed over a range of decreasingly inclusive haplotypic groups. S = the number of segregating sites; $\tau = 2\mu t$; $\theta = 4N_e\mu$, r = Harpending's raggedness index; R_2 = Ramos-Onsins and Rozas' R_2 statistic. Where calculable from coalescent simulations, both the observed value (Obs) and 95% CI (L & U) are provided, in addition to a p-value. Roach s.s. refers to *Rutilus rutilus sensu stricto* and "All" is inclusive of Caspian roach.

Statistic		Grouping					
		All	Roach s.s.	UK Roach	D7	D1	D3
S	Obs	30	17	14	12	10	5
τ	Obs	0.836	0.908	0.645	0.579	0.506	0.293*
	95%L	0.342	0	0.486	0.541	0.381	-
	95%U	1.365	2.123	1.129	1.150	0.984	-
θ	Obs	4.184	2.383	2.026	1.701	1.433	0.930
	95%L	0.697	0.28	0.145	0.142	0	0
	95%U	3.068	1.822	1.447	1.275	1	1.488
	p[θ<=Ob]	0.998	0.998	1	1	0.998	0.890
r	Obs	0.06	0.069	0.101	0.120	0.150	0.184
	95%L	0.026	0.040	0.023	0.053	0.069	0.045
	95%U	0.541	0.902	0.413	0.952	0.987	0.649
	p[r<=Ob]	0.225	0.099	0.533	0.138	0.146	0.607
R_2	Obs	0.029	0.027	0.023	0.025	0.024	0.060
	95%L	0.027	0.022	0.008	0.007	0.017	0.032
	95%U	0.123	0.134	0.173	0.120	0.154	0.178
	p[R^2<=Ob]	0.042	0.050	0.168	0.185	0.055	0.258
Fu's Fs	Obs	-30.87	-26.752	-15.39	-12.902	-12.131	-5.466
	95%L	-14.032	-8.157	-6.900	-7.033	-6.842	-5.976
	95%U	12.758	7.827	5.780	5.295	5.051	4.923
	p[Fs<=Ob]	<0.001	<0.001	<0.001	0.001	0.002	0.026
Tajima's D	Obs	-1.484	-1.445	-1.55	-1.426	-1.406	-0.777
	95%L	-1.534	-1.434	-1.52	-1.354	-1.53	-1.421
	95%U	1.842	2.135	2.24	1.961	2.314	2.149
	p[D<=Ob]	0.03	0.024	0.02	0.138	0.045	0.258

* 95% CI could not be calculated.

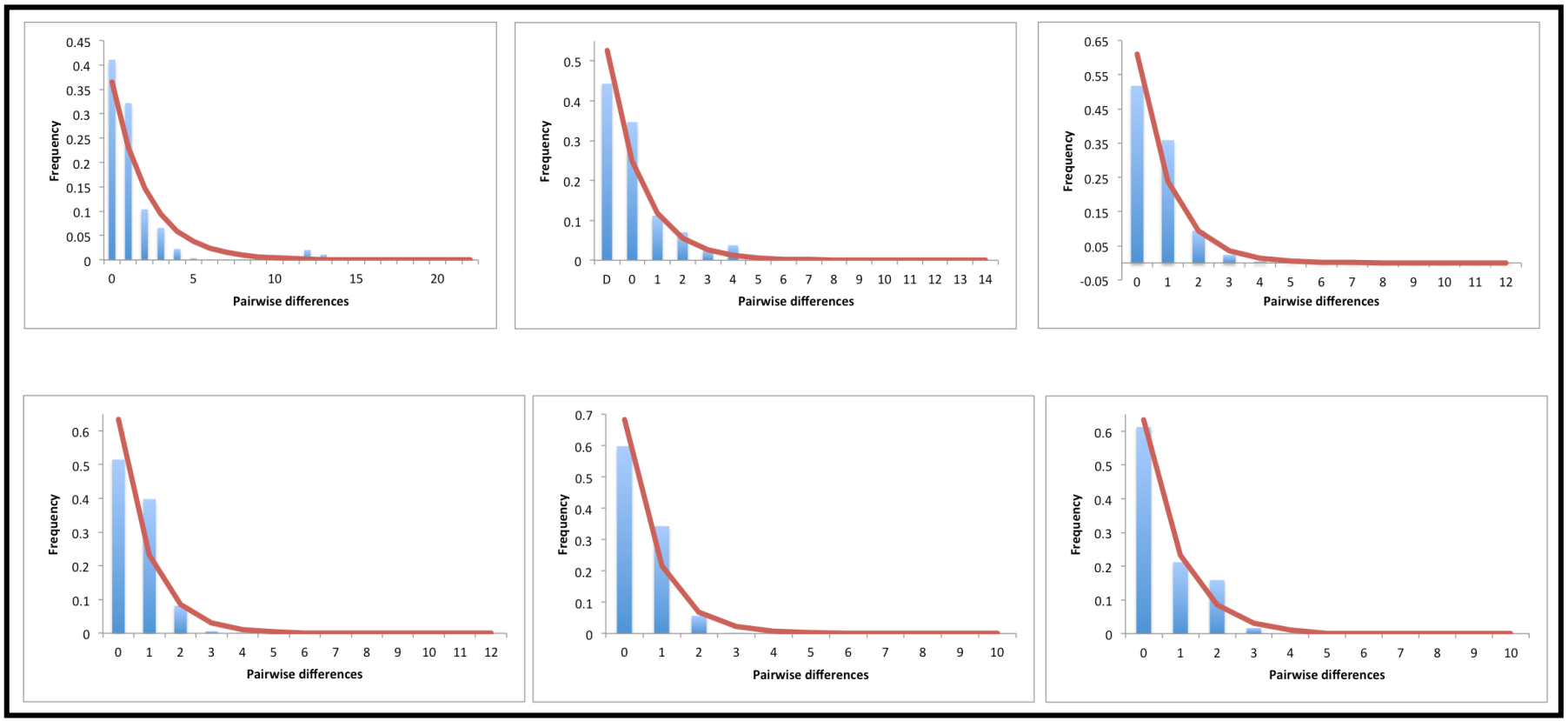


Figure 3.10: Mismatch analyses. Top: from left: All 712 sequenced individuals ("All"), all roach *sensu stricto* individuals ("roach s.s."), 564 individuals from the UK only; Bottom: from left: clade D7 individuals, clade D1 individuals and clade D3 individuals (all based on an MP-derived network). Blue bars give observed frequencies of nucleotide differentiation, whereas the expected number of differences given an expectation of demographic expansion is shown by the red graph.

The distribution of pairwise differences between matrilinearly inherited DNA sequences can be related to the time since the compared sequences last shared a common ancestor (TMRCA) by the equation: μ x p, where p = p-distance (see above) and μ = percent sequence evolution among lineages per million years (Avise et al. 1988; Avise 2000). Fig. 3.11 illustrates the mismatch distributions of nucleotide sequences couched in terms of coalescent ages measured in roach generations ((μ x p)/A_r, where A_r = the age at first reproduction for the roach, 4 years) for all common roaches and for clades D1 and D3. With the exception of D3, shallower coalescences (those occurring within 0-10 thousand generations) are less common than the other categories, indicating that sequences arising from recent mutations are relatively rare compared to ancestral sequences, particularly those sequences that coalesce 20-30 thousand generations ago in the entire roach clade and 10-20 thousand generations ago in clade D1. The shallower clades (D1 & D3), due to their younger age, lack coalescences older than 30 thousand generations, with the majority of coalescences occurring between 10-20 thousand generations years ago. For all roach and the D1 clade, the results are consistent with expectations of past range contraction and expansion resulting in the pruning of the oldest lineages, the prevalence of common, mid-aged haplotypes retained in refugia and a lower frequency of recent coalescences due to the paucity of sampling effort or rarity of new mutations relative to their direct ancestors. That D3 possesses relatively more recent coalescences than the oldest coalescences is probably due to the small number of pairwise comparisons available for this clade (28), compared to D1 (91) and all common roach haplotypes (379) (see Appendix A.3.1 for pairwise p-distances of all common roach haplotypes).

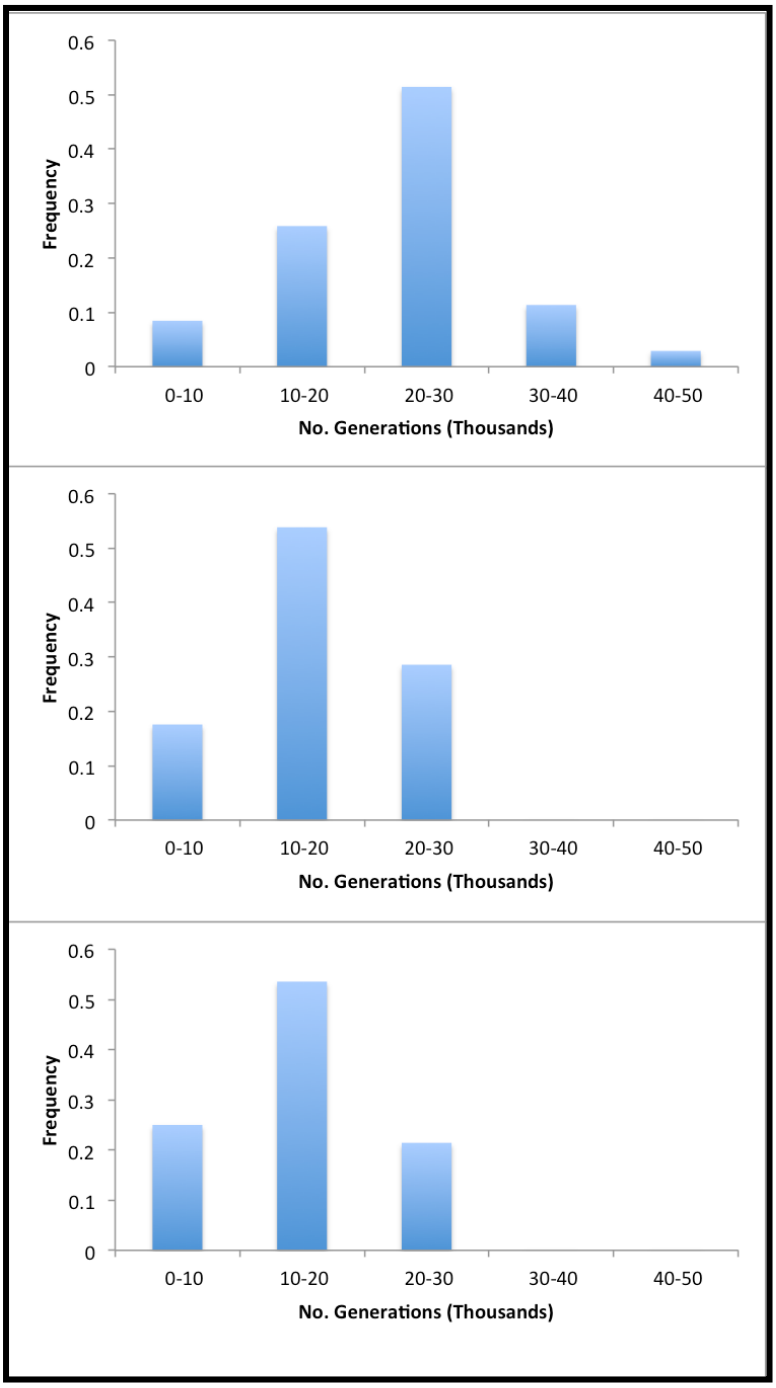


Figure 3.11: Bar charts showing distribution of TMRCA for pairwise sequence comparisons derived from the relationship between p-distance and a substitution rate of 0.0291 substitutions per nucleotide per lineage per million years (see text). Top: All roach with the exception of the Caspian roach; Middle: Clade D1 only; Bottom: Clade D3 only.

3.4 Discussion

3.4.1 mtDNA Diversity

The level of nucleotide diversity observed within the roach for both the cytochrome b locus and for the D-loop segment of the control region are low, consistent with either a selective sweep of the mitochondrial genome in the recent past or due to bottlenecking of the roach population followed by a subsequent demographic re-expansion, likely effected by recent glacial activity (Bell & Andrews 1997; Tipton et al. 2011). In such a scenario most haplotypes are lost in areas that are subsequently made inhabitable (e.g. presence of a massive ice sheet and arctic conditions), and which are then recolonized by surviving lineages from refugia that have been pruned of much of their individual diversity by bottlenecking (Bernatchez & Wilson 1998). Because of the low diversity afforded to river systems, most statistical treatments to delineate pairwise divergent populations (be it F_{ST} or G-statistic) failed, with only a few populations or rivers achieving significance under the conservative application of Bonferroni. Of these rivers, the most divergent were the Witham (by virtue of its fixation of 35 individuals for D1), the Severn, the Kennet and the Sussex Ouse. Most of these rivers were only divergent for D-loop.

The low levels of inter-population divergence are indicative of one or more of the following processes: i) low overall diversity retained from an ancestral population from which all current populations are descended; ii) low mutation rate and/or not enough time has elapsed for mutations to accumulate since the ancestral population(s) became sundered; and iii) there has been ongoing migration, natural or otherwise, since the river systems became isolated and sea levels rose. Each of these factors is not necessarily mutually exclusive from the others. However, the phylogeographic pattern of D-loop variation exhibited by roach in the UK is consistent with Avise's Category V (Avise 2004): in which lineages form part of a shallow gene tree, but some lineages may be widely distributed among river populations (e.g. D1 and D3), whereas others are localized and apomorphic – derived – with respect to the widely distributed ancestral haplotypes (e.g. D2, D5 and D8 are both sympatrically distributed with D3 in the UK and are not found on the European mainland). Overall levels of haplotype and nucleotide diversity of the D-loop (h = 0.482±0.024; π = 0.00101±0.001) are consonant with a recent demographic history incorporating a significant bottleneck followed by rapid population growth, according to the criteria of Grant & Bowen (1998). The above categorizations are

entirely consistent with a scenario of a demographic expansion of roach into the UK following the freeing of the UK's rivers and streams after the last glacial maximum, after suffering population contraction on the continent during the previous glaciation. The mismatch analyses conducted above and further Bayeian inference (see below) provides further analytical evidence.

Freshwater fishes are often cited as having the propensity to develop significant differentiation among populations - in particular those populations that are physically separated by inhabitable coastline and significant expanses of non-floodplain terrain - at least compared to marine species (Ward et al. 1994). Many documented studies of the mtDNA variation of freshwater fishes are consistent with this view, e.g. bulltrout *Salvelinus confluentis* of arboreal North America (Taylor et al. 1998), the fishes *Nannoperca oxleyana* and *Rhadinocentrus ornatus* of the coastal rivers of eastern Australia (Hughes et al. 1999 and Sharma & Hughes 2011, respectively), Scandinavian populations of the Eurasian perch *Perca fluviatilis* (Nesbø et al. 1999), European cyprinids, including populations of the rudd *Scardinius erythrophthalmus* and the vairone *Telestes souffia* in Italy (Stefani et al. 2004 and Salzburger et al. 2003, respectively) and the Iberian species *Anaecypris hispanica* (Alves et al. 2001), *Chondrostoma lusitanicum* (Mesquita et al 2001), *Iberochondrostoma lemmingii* (Lopes-Cunha et al. 2012) and *Rutilus alburnoides* (Alves et al. 1997).

Global surveys of mtDNA variation of European freshwater fish generally reveal low levels of variation throughout their ranges, although particularly so in northern populations, and therefore are in agreement with the expectations of Bernatchez & Wilson (1998) who based their analysis upon an already extensive record of the genetic diversity of North American freshwater fishes. The vimba *Vimba vimba*, although not native to the UK, shares a similar ecology with the roach, in that it is migratory and somewhat tolerant of brackish waters, albeit more so than the roach as it actually reproduces in such waters whereas the roach does not. An analysis of vimba mtDNA from 28 sites located across its range revealed only 17 D-loop haplotypes (from 86 individuals and 500bp of sequence) and 15 cytochrome b haplotypes (from 44 individuals and 600bp). The concatenated data yielded 21 unique haplotypes (Hänfling et al. 2009). For both the D-loop and cytochrome b, only 1.3% of all sites were polymorphic. The respective percentage values for the UK roach are 2.2% and 1.17%. Another fish with a similar ecology to that of the vimba that has been extensively studied at the level of genetic diversity is the Eurasian perch. An extensive study of D-loop variation from 56 distinct populations spanning the west of

Europe to the western reaches of Siberia uncovered 35 haplotypes from a 365bp survey of 488 individuals (Refseth & Nesbø 1998; Nesbø et al. 1999). The 21 variable sites responsible for the haplotypic diversity represent some 5.75% of the D-loop sequence. A pan-European survey of cytochrome b variation of individuals from much of the roach's European range (not including much of its northern distribution, i.e. the UK, Scandinavia and northeast Europe), with a strong emphasis on roach inhabiting the Meuse and Scheldt drainages of Belgium, found that a 475bp fragment from 265 individuals sampled from 52 locales described 70 unique haplotypes based upon 77 variable sites (16.21%) (Larmuseau et al. 2009). From a re-analysis of Larmuseau's data inclusive of the UK roach, it is revealed that UK roach possess unique haplotypes, thus increasing the amount of diversity present within European roach by a small degree (six haplotypes).

Most studies of mtDNA variation in widely distributed European freshwater fishes have a broad sampling scheme, contingent upon a primary motivation of deciphering the postglacial and interglacial histories of species, although there are exceptions (e.g. both the Aegean (Gollman et al. 1997; Imsiridou et al. 1997, 1998; Durand et al. 1999; Zardoya et al. 1999; and Tsipas et al. 2009) and Italian regions (e.g. Stefani et al. 2004; Salzburger et al. 2003) have been well studied with respect to their localized cyprinid fauna**).** Due to the compromise between sample sizes per location and the extent of sampling required over a species' range necessitated by restricted timeframes and limited budgets, often only a small sample of regional mtDNA diversity is generally described. Within the UK, only very few rivers have been assessed for mtDNA diversity across potamodromous freshwater fishes. Nesbø et al.'s study of perch (1999) found that all ten individuals taken from the River Thames were monomorphic for the D-loop haplotype, 'F'. The inclusion of four tench *Tinca tinca* from two private ponds located within the floodplain of the Yorkshire Ouse uncovered two distinct cytochrome b haplotypes (3 individuals of an 'eastern European' haplotype and one of the 'western European' haplotypes) (Lajbner et al. 2011) sharing some aspects of haplotypic distribution observed in roach within the same area (see below).

The common bream *Abramis brama* is often co-distributed with roach in lowland waters. A study of cytochrome b diversity within bream (Hayden et al. 2011) utilized 25 individuals from five rivers (5 bream per river): River Hunstpill (artificial), River Ouzel (Great Ouse watershed), River Rother (Arun watershed in West Sussex) and the Rivers Bourne and Kennet (both within the Thames). The sequence analysis of a 638bp fragment revealed that all 25 bream were monomorphic except for one individual from the Rother

that closely matched cytochrome b sequences from the related white bream *Blicca bjoerkna* and is present either due to past introgression with the common bream or through sampling misidentification.

Nine chub *Squalius cephalus* surveyed within the Thames were monomorphic for the same 600bp 'Atlantic' cytochrome b haplotype commonly found elsewhere in western European populations of this fish (Durand et al. 1999). 25 dace *Leuciscus leuciscus* individuals sequenced for 715bp of the cytochrome b locus found some variation within and between two populations located in the Rivers Frome and Allen in southwestern England (Costedoat et al. 2006). The River Allen's 12 dace included four unique haplotypes (H1, 3, 7 & 8), whereas the 13 dace from the Frome had 3 (H1, 2 & 4). It is clear that at least in the southwest, dace populations are more variable than roach populations in other parts of the UK at the cytochrome b locus (roach: mean π = 0.0005±0.0002 and mean h = 0.1983±0.0689; dace: mean π = 0.0013±0.0012 and mean h = 0.5612±0.1292). Any difference needs to be qualified by the size of fragments analyzed and the location and history (e.g. native or introduced) of any particular location.[19]

Only one in-depth study of mitochondrial sequence data for a widely distributed fish sampled from across its UK distribution has been published to date, Dawnay et al.'s (2011) comprehensive coverage of grayling *Thymallus thymallus* populations. A major component of this paper assessed D-loop diversity (825bp fragment) for 5 individuals each from 27 distinct populations. The scope of this study oversaw D-loop diversity located within 17 distinct drainages, although the overlap of river systems with those covered by the present study is restricted to just six catchments: Thames (River Kennet), Trent (River Derwent), Hampshire Avon, Yorkshire Ouse (River Ure), the Severn and the River Test. 85 sequenced individuals yielded 10 unique haplotypes from 7 variable sites. The grayling study shares some parallels with the above assessment of roach D-loop diversity: most notably the prevalence of a single haplotype in the majority of populations. The H1 haplotype in grayling is found in all but two of the 27 surveyed populations, whereas the D1 haplotype of the roach is found in all of the surveyed populations. However, unlike the roach, grayling were monomorphic in over 50% of surveyed sites (15/27, 13 of which contained only H1). The two monomorphic populations of roach were also identical for the most common haplotype. Interestingly, the Severn population of grayling is composed entirely of unique haplotypes (H4 & H7), whereas the Severn population of roach is

[19] The roach is thought not to be native to southwestern England, which may also be the case for the dace and other coarse fishes.

divergent from the other river populations in that D1 is reduced to minority status and D3 is much more frequent in occurrence. The River Kennet population of grayling, however, is amongst the uniform collection of H1 populations, whereas it is a locale containing much haplotypic diversity in roach.

The sum total of 85 sequenced grayling yielded an average of one unique haplotype per 8.5 grayling sampled. The equivalent figure for the UK roach population, based upon a lesser fragment size of 634bp derived from 564 roach, is 35.25. Undoubtedly, an important component of this shortfall in diversity is the shorter sequence length used in the roach analysis. However, the lower diversity observed in the roach compared to the grayling may stem from three alternate identifiable sources: i) that present day descendants of roach were isolated more recently than those of the grayling; ii) the D-loop mutation rate is lower in the roach than in the grayling; and iii) the roach possesses a lower evolutionary effective size than the grayling (but see Taylor et al. 1998). Of these three possibilities, the latter is the least likely. Previous studies of roach have identified extensive within-population variability – and low inter-population divergence -at the level of local populations, at least in nuclear genomes, indicative of high effective sizes even in the face of past demographic instability (e.g. Bouvet et al. 1991, 1995; Baranyi et al. 1997; Wolter 1999; Hänfling et al. 2004; Demandt & Björkland 2007; Demandt 2010; and see Chapter Four). Additionally, the roach is widespread and extremely common, and thus liable to posses much more intrinsic genetic variation as a result of high census numbers (McCusker & Bentzen 2010), making it unlikely to suffer a paroxysmal loss of genetic diversity throughout its range, although peripatric populations or those on the fringes of a geographical distribution may experience significant loss of diversity even in the face of regenerative migration from the core of a species' distribution (Johannesson & André 2005). Furthermore, the grayling is a regionally threatened species that has suffered a number of population bottlenecks in much of its wider distribution (e.g. Swatdipong et al. 2009), more so than the more common coarse fish of the UK, and is hypothesized to account for many of the fixed populations found in Dawnay et al.'s study (further backed up by microsatellite data indicating significant F_{ST} scores among drainages).

It would be reasonable to expect less D-loop variation within grayling populations than roach populations. Therefore, it may be that the nucleotide substitution rate differs between these two species, or that their recent demographic histories are asynchronous. By assessing the evolutionary relationships between haplotypes and their geographical and demographic properties, the underlying causes for the slight incongruence in D-loop

diversity between the roach and the grayling may be revealed. However, the similarities in haplotypic distribution and overall low levels of diversity are marked, suggesting that any differences are of degree, not in kind, and that the broad glacial landscape of the last epoch is chiefly responsible for purging the UK of much, if not all, of its eurytopic freshwater fish fauna, thereby further validating the boreal genetic depauperation paradigm of Bernatchez & Wilson (1998).

3.4.2 Population Structuring

Roach populations within the UK show little divergence on the basis of the distribution of unique haplotypes among drainages and zero concordance of geography with monophyletic mtDNA lineages. Similar patterns have been found in a number of freshwater fishes. Like the roach, the perch shows similar low levels of diversity in a northern expanse of its distribution: Refseth et al. (1998) found that 109 individuals sequenced for 378bp of the D-loop from 20 sites scattered within Norway and Sweden yielded 12 haplotypes based upon 10 variable sites (2.65%). Such low within-region diversity, especially acute in northern populations, or areas that have only recently been recolonized, will translate into low levels of inter-population divergence, particularly among drainages, with the potential for high levels of within-drainage diversity in hierarchical analyses of structuring. The application of hierarchical Φ-statistics did confirm that variation among roach populations was present among watersheds (5.29%), at least for the D-loop, albeit unsupported statistically. However, 7% of the variation exhibited across the hierarchical scale could be attributed to diversity within rivers. In fact many of the rivers in which samples were taken at two sites within a single stretch exhibited a range of differentiation (e.g. F_{ST} within the Yare, the Suffolk Stour and the Yorkshire Ouse is 0.555, 0.133 and 0.086 respectively). Despite there existing differentiation among populations and some rivers by virtue of the variation of haplotype frequencies among them, no clear association between individual haplotypes or ancestor-descendent relationship can be ascribed to any particular drainage in the UK except for the Severn and Sussex Ouse (in terms of low frequencies of D1 and elevated frequencies of less common haplotypes) and the Thames, in which five direct descendants are located (four haplotypes are descended from D1, including D3, and two from D3 itself), and the Great Ouse (five: three haplotypes derived from D1 including D3, and two from D3 itself). The UK roach harbor both derived and ancestral haplotypes, from within the common roach lineage, including D4 from which all non-Caspian haplotypes are inferred to have

derived, with the exception of D37 (Fig 3.8). A secondary lineage, derived from D9, is not represented in the UK (except for three D9 individuals from the Great Ouse), but appears to be a sub-sample of a much greater, and potentially much more diverse European lineage not fully investigated by the limited sampling presented here.

Superficial similarities with roach in the distribution of mtDNA variation have been observed in some obligate freshwater fishes in distant locales that have also recently been perturbed by Pleistocene climatic upheaval, for example, *Hypseleotris compressa,* from eastern Australia, revealed very little differentiation among 15 coastal river systems for a 567bp fragment of the ATPase 6 locus, although differentiation was also lacking within watersheds (McGlashan & Hughes 2001). Using Wright's hierarchical *F*-statistics, the authors report that 98% of all genetic variation was found within populations. A survey of an 865bp D-loop fragment in the bleeker *Leptobotia elongate* within China's upper Yangtze River found little structuring among groups (tributaries) (1.7% of total variance), with the vast majority of haplotypic variation found within individual populations (Liu et al. 2012). Whilst not as high as that found in the roach and in the bleeker, a study of a 715bp cytochrome b dataset of variation in the dace from across central and western Europe apportioned 69.22% of all variation to that found within populations (with 17.17% apportioned to differences between populations and 13.01% to differences between drainages) (Costedoat et al. 2006). A further complication of assessing population structuring in areas with complex glacial histories is that AMOVA analyses are prone to underestimate the extent of structuring when populations have a historical legacy of range contraction and expansion (Templeton 1998). Sampling a range of watersheds with variant glacial histories may explain the greater structuring found among European dace for cytochrome b variation than higher resolution studies utilizing the D-loop. These findings are consistent with the characteristic of low mtDNA variation due to repeated glacial purges of northern latitudes and the limited temporal scope for mutations to accrue in different refugial populations whose inhabitants, whilst demographically allopatric, still closely resemble one another in their mitochondrial nucleotide sequences.

MtDNA diversity exists within UK watersheds, particularly D-loop haplotypic diversity, however limited in extent. Scientific consensus predicts that in riverine systems with pronounced downstream river flow, it is expected that a parallel pattern of asymmetric downstream-biased dispersal of larval and juvenile forms of aquatic species would be observed, such that more haplotypes (and greater quantitative measures of genetic diversity) are found downstream as opposed to sites situated upstream (e.g. the razorback

sucker *Xyrauchen texanus* of Western North America (Dowling et al. 1996)). Examples within the freshwater fish literature of the *a priori* expectation of decreasing diversity with upstream distance are most commonly found in species in which a source-sink metapopulation dynamic is most pronounced. For example, many upstream populations in the endangered cyprinid *Anaecypris hispanica* are fixed for mitochondrial genotypes, probably because upstream populations are often isolated and extirpated by regular periodic droughts (Mesquita et al. 2001). Likewise, similar source-sink dynamics were described for the demersal bullhead *Cottus gobio*, a species characterized with low motility, within a tributary of the River Rye, UK, that also displays a pattern of significant isolation-by-distance (Hänfling & Weetman 2006). For other species, particularly vagile species, there is often found a lack of differentiation and structuring among sites within rivers (e.g. Li et al. 2012).

These *a priori* expectations of an increase of mitochondrial diversity with distance downstream were not matched within the roach. Diversity was greater upstream than downstream in the Yorkshire Ouse (two haplotypes are found downstream at Beningborough, whereas four haplotypes are observed in the upper tributary, the River Ure; YOC: $\pi = 0.500\pm0.135$, $h = 0.036\pm0.0013$; YOB: $\pi = 0.111\pm0.096$, $h = 0.003\pm0.0032$), the Great Ouse (seven haplotypes are found in the upper Ouse at Newport Pagnell compared to the four found in the middle Ouse at St. Ives; GON: $\pi = 0.593\pm0.100$, $h = 0.040\pm0.088$; GOE: $\pi = 0.383\pm0.120$ $h = 0.013\pm0.047$) and the Suffolk Stour (three upstream haplotypes at Stoke-by-Clare compared to just the ubiquitous D1 haplotype downstream at Higham; SSS: $\pi = 0.385\pm0.102$, $h = 0.013\pm0.038$). In the Trent, the River Derwent ($\pi = 0.0044\pm0.0037$, $h = 0.133\pm0.112$), situated more closely to the downstream confluence with the main Trent than either the Mease of Sence, is less diverse than either of the more upstream located sites in the Rivers Mease ($\pi = 0.013\pm0.0055$, $h = 0.389\pm0.164$) or Sence ($\pi = 0.017\pm0.0034$, $h = 0.509\pm0.101$). Only in the Thames and in the Medway was this situation not observed. A tri-locus study of mtDNA variation of the flannelmouth sucker *Catostomus latipinnis* within the huge expanse of the Colorado River, USA, uncovered a clinal distribution of nucleotide diversities that increased with distance upstream (Douglas et al. 2003). Unlike the roach, however, haplotype diversities were high at all locations.

Dispersal ecology in any study of genetic variation and population connectivity is an important element that will impact upon expected results (e.g. see Blasco-Costa et al.'s study of the effect of vector organism dispersal behaviour upon levels of within-stream

genetic diversity of two endoparasitic organisms (2011)). For instance, upstream populations experience less immigration than those located further downstream (due to a reduced pool of migrants from upstream headwaters and less migrants originating downstream due to the vigor of swimming against the flow). Therefore, novel mutations that occur upstream, or are present through *ex situ* translocation, are likely to persist for a greater period of time than novel variation located downstream where increased immigration relative to emigration decreases the likelihood of any single variant approaching detectable frequencies (Nagylaki 1978)[20]. This insight into asymmetric dispersal ecology has only recently been developed within a theoretically robust framework for a range of applicable habitats (e.g. coastal fauna restricted to tidal zones (Pringle et al. 2011)). Such processes, as well as strictly limited unidimensional gene flow, will promote the formation of genetic clines within such environments. However, the sampling scheme adopted in this study was not set-up to detect clinal change within linear stretches of river and so must remain unknown until further study.

3.4.3 D-Loop Phylogeny and Demography of the UK Roach Lineage

The mean estimated *Rutilus rutilus/caspicus* D-loop divergence rate of 1.46% per million years is lower than the general mitochondrial mutation rate of 2% divergence per million years often assumed for bony fishes, but within the range of low-end mutation rates for the D-loop within some teleost fishes (which can be as low as 0.5% (Cantatore et al. 1994) and 0.8% in the brown trout *Salmo trutta* (Osinov & Bernatchez 1996)). Engelbrecht et al. (2000) reported that the D-loop diversity in European populations of the bullhead *Cottus gobio* was probably the result of a mutation rate lower than the initial 2% sequence divergence per million years assumed for their analyses. Reduced mutability of the mitochondrial genome as a whole compared to other fishes was found in the sturgeon and paddlefish family, the Acipenseriformes (Krieger & Fuerst 2002), and in the D-loop in particular among species of killifish *Fundulus* spp (Whitehead 2009). The mutation rate of mitochondrial DNA as a whole in fishes has often been reported as being slower in this taxon than in other vertebrates, perhaps by as much as five-fold (Martin et al. 1992).

[20] In Nagylaki's words: "If dispersion is preferentially out of an environmental pocket at the end of a very long habitat, the condition for maintaining the allele favored in the pocket becomes less stringent than for symmetric migration; dispersion preferentially into the pocket increases the severity of the condition for polymorphism." (1978).

However, within the widespread western European roach clade the inferred mutation rate was much higher: a high of 7% pairwise divergence per million years with a mean value of 5.82%. The 2-3 fold increase in rates estimated for more recent divergences is likely due to the over-saturation of ancestral sequences, whereby mutational events are superimposed upon more ancient point mutations, thereby resulting in a dramatic underestimation of mutation rate. Therefore, much less confidence should be attributed to rates inferred from evolutionary distant bifurcations. Moreover, the estimate of mutation rate derived from the roach data is a product of an underlying mutational model (estimated here using a ML criterion) and a limited, somewhat biased sample of fish used to construct an estimated phylogeny that may or may not accurately portray the true phylogeny. Additionally, demographic processes may affect the estimate of mutation rate. For instance, past population contractions remove much of the available variation from subsequent observation, thereby underestimating the true rate of mutation (Tajima 1989)[21]. Therefore all estimates of divergences and demographic expansions based upon the rate estimated at the point that *R.rutilus* and *R.r.rutilus* diverged must be applied extremely cautiously.

From a cursory observation of Fig 3.8, the distal haplotypes within the lineage observed from UK waters are both commonplace and numerous with a surfeit of recently evolved, low frequency descendants, forming a 'star' pattern. Such a pattern is indicative of recent demographic expansion. From the mismatch analyses of the D-loop, and applying nucleotide substitution rates inferred from the common roach clade only, the origin of the common roach clade was dated at 20-25 thousand years before the present, placing the lineage diversification and expansion in the upper-Tarantian stage of the late Pleistocene epoch. The major sub-lineage found within UK waters is D1, from which D3 is derived. D1 is estimated to have derived from a range expansion dating from 14.5 – 17.5 thousand years ago, whereas D3 is dated at around 7-8 thousand years ago, the approximate date at which the UK became sundered from mainland Europe by the creation of the English Channel. However, application of the lower substitution rate does not support the consonant geologic and demographic events above, but this may be due to the limitations elaborated previously.

The application of mismatch analyses to infer past demographic processes is only applicable when the number of superimposed nucleotide substitutions is limited or zero (Schneider & Excoffier 1999). However, the problem of nucleotide site substitution

[21] Mutation rates based upon mutations at low frequency are less affected by past demographic processes because they are likely to be genealogically younger (Kimura and Ohta 1973).

saturation is only a problem when the per site mutation rate is so high - or that the expansion happened so long ago - for saturation to have accrued. In the present case pairwise estimates of distance are invariant regardless of various weightings given to transversions over transitions, or of varying nucleotide frequencies. The most simple distance, the p-distance, suffices to describe differences among haplotypes, therefore the mismatch analyses are based on differences likely to have accrued only once per site and estimates of divergences from mismatch parameters retain theoretical validity.

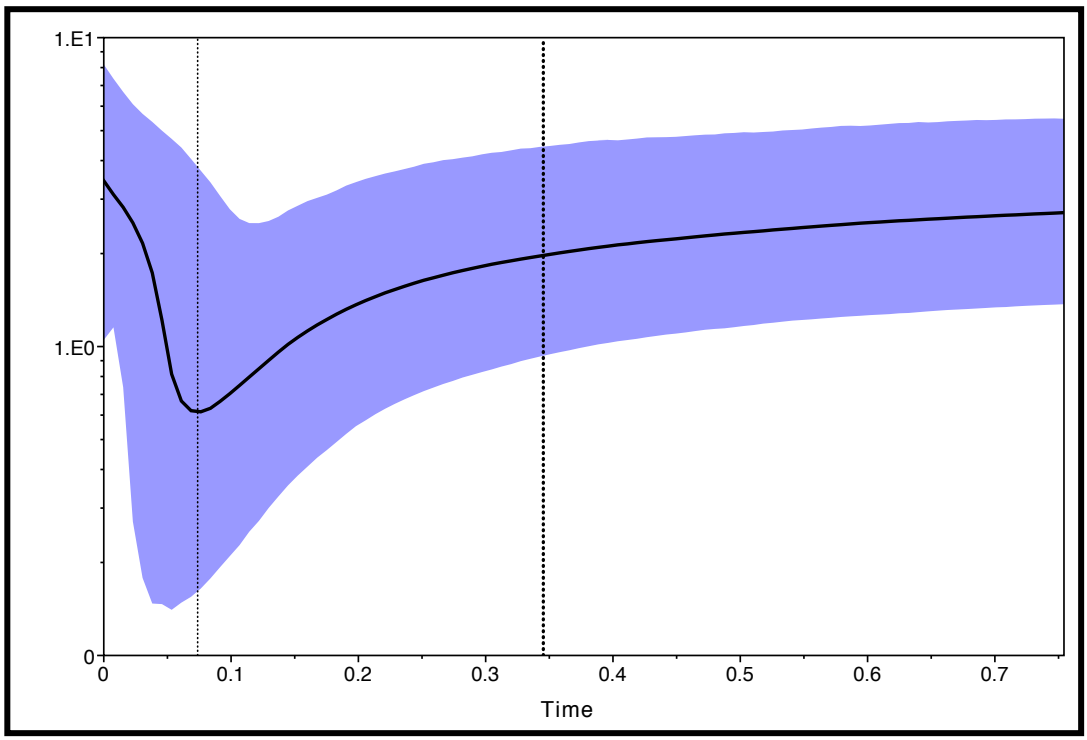


Figure 3.12: Skyline graph depicting the temporal changes in female effective population sizes over time (N_e(f), y-axis). Bold dashed line = Median date of divergence from the lineage leading to the Caspian roach; normal dashed line = lower bound of 95% HPD. The blue shaded area represents the 95% HPD of effective population size estimated over time (millions of years).

According to the graph depicted in Fig 3.12, there was a gradual reduction in common roach effective population size that accelerated during the mid-Tarantian to reach a nadir of diversity approximately 75000 years ago, coinciding with the beginning of the Würm glaciation (70000 – 10000 years ago). The y-axis in Fig 3.12 plots a composite effective population size (whereby the log scale represents a population size described by $N_e*\tau$, where τ = number of generations, if the substitution rate to calibrate the internal phylogenetic clock was stipulated per year[22], but is presented here as the number of

[22] However, here τ describes the number of generations elapsed in a million years – given a generation time of four years - as the mutation rate was specified per million years in the BEAST analysis.

generations per million years (around 250000)). At the median point of the origination of the common roach, $N_{e(f)}$ is estimated to have been around 275000 individuals, which plummeted to less than one twentieth of that number (approx. 12500) during the nadir before rebounding to a current estimate of around 312000 individuals. However, unless all genetic diversity, or the entire range of a species is sampled, all estimates of effective size will tend to underestimate the theoretical maximum number of contributing individuals.

To further illustrate the demographic influence of range expansion upon mtDNA variation, consider Fig 3.13 where the number of lineages accrued over time is plotted on a logarithmic scale against time (forwards in terms of bifurcations/backwards in coalescences). Nee et al. (1996) demonstrated that for neutral loci the expected shape of the graph for a population undergoing demographic expansion is one that tends to asymptote horizontally towards the present, indicating a relative paucity of new lineages compared to more ancestral nodes in the tree. The general shape of the graph in Fig 3.13 is as described above, however more recent times have seen an increase in branching events that could be indicative of relative demographic stability, or the increase in lineages is a statistical artifact.

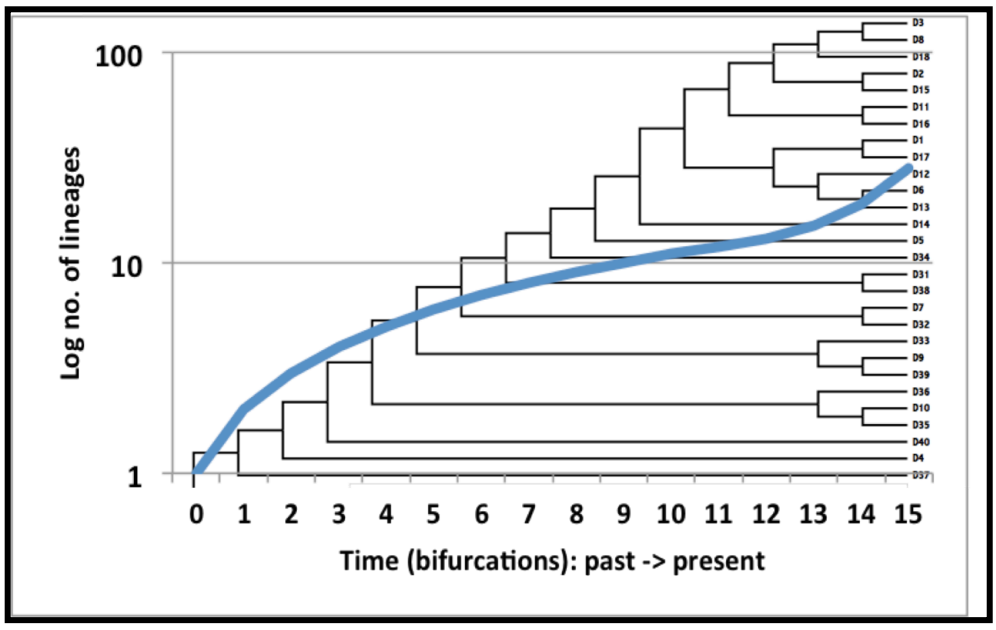


Figure 3.13: Lineages through time plot (logarithmic scale on y-axis) showing the increase in lineages with time from an ancestral haplotype (blue graph) superimposed upon a topographical representation of the roach phylogeny as inferred by the ML approach.

The Würm glaciation was characterized by periodic warming events known as Dansgaard-Oescgher epsiodes (interstadials). As a result of glaciation, sea levels around the world

began to decline around 50kya before falling sharply at around 30Kya to the levels found at the LGM (around 125m below present day levels (Chappell 1974)). Mean cooling in western Europe was in the region of 10^{o}c from the period 85 – 30kya (Genty et al. 2003). The last stadial period was a particularly severe Heinrich event [23](H5a) (Hemming 2004), sandwiched between two cooler interstadials. Range contractions resulting from these glacial fluxes would result in the bottlenecking of lineages that originated from before this time. This scenario has been induced to explain significant genetic upheaval in Holarctic freshwater fish populations (e.g. the North American blue catfish *Ictalurua furcatus* (Pradhi 2011)). However, interstadials are generally more calm climatic periods and may have presented the opportunity for biotic expansion and re-diversification before experiencing range contraction once again. This geologically recent period of severe climatic flux may explain why so few branches survive from the more ancestral haplotypes.

However, since 'bottoming out', the data suggest that the roach has undergone rapid demographic growth in the ensuing millennia, perhaps as a result of inhabiting significant refuges in southern and southeastern Europe, both in terms of effective population sizes (a minimum of 12500 effective females is posited here), and perhaps in number (nominal locations include the lower Danube, the Black Sea and southern and western Europe). The blue catfish is inferred – from D-loop data – to have been reduced to 400000 effective females from a pre-decline size of some 1.2 million (Pradhi 2011). The discrepancy with roach, in terms of sizes, probably reflects the limited scope of sampling of the common roach for D-Loop sequences, and these size estimates, along with other parameters inferred from BEAST, may under-estimate the true values. Even so, the expansion of 'Clade D3' may be consistent with an expansion of roach following the retreat of the Devensian glacier from the British Isles and the subsequent sundering of the islands from the European mainland.

The majority of mtDNA variation within the UK distribution of roach is found in the southeastern and southern areas of England. This scenario is expected if this area, which once drained into the huge Channel River system prior to the mass denudation of 7500 years ago, received far more waves of colonizing roach, than areas further away from the main European areas of refuge. However, the two most northerly rivers assayed for roach (the Tees and Yorkshire Ouse) contained within them either high levels of haplotypic and

[23] Heinrich events are associated with catastrophic ice breaks and extensive rafting of icebergs in the polar and Holarctic regions.

nucleotide diversity (Tees) and/or ancestral haplotypes found in mainland Europe and not, or rarely found, within the UK (see Figs 3.2, 3.3 and 3.4). Haplotypes D16 and D7 are both found in the Tees, southern UK and mainland Europe, whereas the most ancestral UK haplotype, D4, is found in the UK only within the Yorkshire Ouse. The latter two haplotypes are ancestral and the temporal placing of D16 doubtful. Three explanations can explain these results: first, that there has been sufficient movement of roach across Europe, or from southern England, to have mixed lineages within individual drainages and transported the more ancestral haplotypes to the north and northeast; second, the presence of ancestral haplotypes in the north and northeast were deposited by a second, northerly wave of colonization after the last ice age; third, the presence of ancestral haplotypes in the north is a relic signal of the first wave of colonists, which were subsumed in more southerly locations by a second wave containing more diversity. There is no direct evidence for the former explanation, apart from a long and mostly anecdotal litany of instances of regional and localized fish movements, at least until the 20th century. However, the pattern of D-loop diversity observed within the UK population of the grayling is similar to that observed in the roach, whereby most diversity was observed in southern and northeastern drainages, including a number of basal and European haplotypes in the north (Dawnay et al. 2011). This parallel patterning of D-loop diversity is consistent with the second and third explanations regarding colonizing waves into the UK, at least for roach and grayling.

3.4.4 mtDNA as a Utilitarian Marker of Differentiation

The D-loop is a common genetic marker in population genetic and demographic studies because it is widely found to be the most variable and mutable of genetic markers, making it a very useful tool for use in intraspecific studies of species with shallow phylogenetic origins. The locus has been shown to possess a mutation rate up to an order of magnitude greater than nuclear coding sequences within fishes (Meyer 1993). D-loop variability has been found within teleosts, including the cyprinids (e.g. the rosy bitterling *Rhodeus ocellatus* **(**Kawamura et al. 2001), *Barbus* spp (Dimmick et al. 2001), the *Distoechodon* genus (Liu 2002) and the vimba (Hänfling et al. 2009).

In the present study of roach, although the mean number of screened individuals per unique haplotype was lower for the cytochrome b locus (425bp) over the D-loop (634bp), the D-loop uncovered a more complex genealogy than the cytochrome b survey, and

almost all of the 98 roach surveyed for cytb were bearers of the most common haplotype (89 (90.8%) as opposed to just 396/564 (70.2%) of roach screened for the D-loop). Contrast these data with the limited set of concatenated sequences in which 30/49 roach bore CON1 (61.2%). There is a clear positive relationship between an increase in sequence information with sampling proficiency *and* sampling range. However, sampling effectiveness does not only depend upon sampling all demes or populations, which may be a waste of effort, but also depends upon recent demographic histories and coalescent times (Wakeley 2008). Therefore, it is perhaps more prudent to increase the information content of the locus in question, rather than increase sampling density or proficiency at any particular site. With this in mind, almost any variable marker will have some utility at some temporal or spatial level of analysis. For example, a small fragment of the cytochrome b locus (307bp) was proven highly informative in assessing population structuring within and among populations of the economically important percid *Channa marulis* of Indian rivers systems (Habib et al. 2011).

One major complication of utilizing mtDNA as a genetic marker is that it may not be a suitably neutral marker – that is to say that the distribution of haplotypes may be influenced by natural selection. If so, this may impact all studies of phylogeography from which inferences are based upon an assumption of strict neutrality. Ballard and colleagues (2007) found that mtDNA sequence variation in genes coding for subunits of the cytochrome oxidase c protein co-varied in response to life-history traits (egg size, fecundity, etc) within the fruit fly *Drosophila simulans*. A similar linkage of mtDNA selection with life-history characteristics was observed for the cytochrome b locus in fringillid songbirds (Rottenberg 2007). This study found a close association between increased longevity and cytochrome b evolution that minimized the production of senescence-inducing free radical species of oxygen. Variation of the NADH6 locus in domestic horses *Equus caballus* was found to vary with altitude, with both directional and non-directional selection implicated (Ning et al. 2010). Altitude has also apparently played a significant role in apportioning variation in mtDNA sequences within a wide range number of species, e.g. cytochrome b variation among *Peromyscus* deer mice populations of varying altitudes is largely determined by purifying selection (Gering et al. 2009).

Two of the cytochrome b haplotypes found within the UK roach population each have a single, distinct non-synonymous nucleotide change: “UK” – an isoleucine has substituted a valine residue; and “KenSo” – an alanine has replaced a threonine. The former substitution

may have a severe impact[24] upon the equilibrium constant of the cytochrome b molecule, which may affect mitochondrial efficiency (George & Blieck 2011). Multiple authors have postulated adaptive changes within the cytochrome b locus for a range of taxa, including freshwater fishes (Moyer et al. 2005). For a great many of these studies, climate is linked to adaptive functionality of the mitochondrial loci under study (e.g. Ruiz-Pesini 2004). Foote et al.'s 2010 study of whole mitochondrial genome diversity within and among pods of killer whales (*Orcina orca*) also uncovered significant non-synonymous amino acid changes in each of two different ecotypes, suggestive of variant functionality. Ecological co-variance with mtDNA haplotype was also found among populations of *Anolis* lizards inhabiting similar ecotypes in the Lesser Antilles (Malhotra & Thorpe 1994).

However, many of these studies in which directional selection was inferred among species may be alternately explained by a combination of structural constraints and purifying selection (Ingman & Gyllensten 2007). Bazin & colleagues (2006) noted the non-appearance of a population size effect when the within-species variation of nuclear and mitochondrial sequences were compared among 1600+ species of animal, explaining this disparity in variation (average mtDNA diversity was similar across taxa, whereas nDNA diversity was greater in abundant taxa) by invocation of widespread selective sweeps of mitochondrial genomes. It is becoming increasingly clear, therefore, that functional genes within the mitochondrial genome may be under significant selection pressure - or have been so in the recent or distant past - such that an imprint of selection may be present within extant patterns of haplotype distribution, perhaps superimposed upon similar patterns derived from other sources (e.g. demographic history) (Galtier et al. 2009).

The D-loop may also be influenced by selection, either directly or through genetic hitchhiking (genetic draft), by virtue of its linkage to all other coding genes in the mitochondrial genome and an absence of mitochondrial recombination. D-loop diversity has been observed to co-vary with environmental factors in some published studies, e.g. D-loop diversity in the white-toothed shrew *Crocidura russula* declines with altitude (Ehinger et al. 2002). In this study, a regression of D-loop and cytochrome b diversities with elevation above sea level for each of the sampling sites (Fig 3.14) suggest that there is some non-random relationship between D-loop haplotypic diversity and altitude (r = 0.360, 2-tailed p-value = 0.065), but none whatsoever for the cytochrome b data. The relationship is also present, but very slight and insignificant, for nucleotide diversity as

[24] $P < 0.01$ that the mutation occurred through chance alone (from a cursory analysis of a shallow cytochrome b phylogeny within the roaches, implementing a ML approach to detecting significant non-synonymous amino acid changes in TreeSaap (Wooley et al. 2003)).

well (r = 0.236, 2-tailed p-value = 0.236). These data alone cannot differentiate among the following hypotheses that may account for the increase in D-loop diversity with altitude: 1) Different D-loop haplotypes are increasing in frequency due to genetic draft whereupon the D-loop is hitchhiking with another functional gene. The nature of D-loop mutation rates may inevitably lead to different D-loop haplotypes being drafted for the same mtDNA functional mutation; 2) D-loop haplotypes - or their linked neighboring loci - are part of an ongoing frequency-dependent process; 3) The increase in diversity with altitude is due to an increase in environmental heterogeneity with distance upstream, whose populations are more susceptible to bottlenecks and lineage sorting of mitochondrial genomes; 4) Non-equilibrium effects due to a relative reduction of immigration in upstream populations enabling mutations to persist when asymmetric migration rates are high; or, 5) Life-history effects of differential age-structured populations.

It has been suggested that roach in the upper reaches of UK streams are larger, older and fewer in number than in populations located more downstream (S. Axford pers comm). Female roach do show variation in life-history characteristics, at least in differing climates in Norway (Vøllestad & L'Abée-Lund 1990). Those roach found furthest north grew more slowly (5.6mm yr^{-1} in the most northerly site) with lowest fecundity and lowest gonadosomatic index (GSI) and the production of larger eggs, although no increase in body size was observed (in fact the reverse – an increase in fecundity with length – was recorded). The authors suggest that the data support a bet-hedging model of life-history evolution. Similar expectations may be met for roach in the upper reaches of rivers. Following on from Gillespie's analysis of reproductive variance and fitness (1974), Shpak (2005) showed that for metapopulations with few demes and an island model of population structure, a reproductive strategy in which both the mean and the variance of offspring production is low relative to the rest of the population would be favored in demes that experience little immigration.

It must be noted that many of the above hypotheses are not necessarily mutually exclusive and may be acting in tandem. However, traditional experimental, field and further genetic analyses may test hypotheses 4 and 5. Moreover, more extensive geographic surveys conducted with multiple nuclear and mitochondrial loci at greater sequence/locus resolution may untangle hypotheses 1-3.

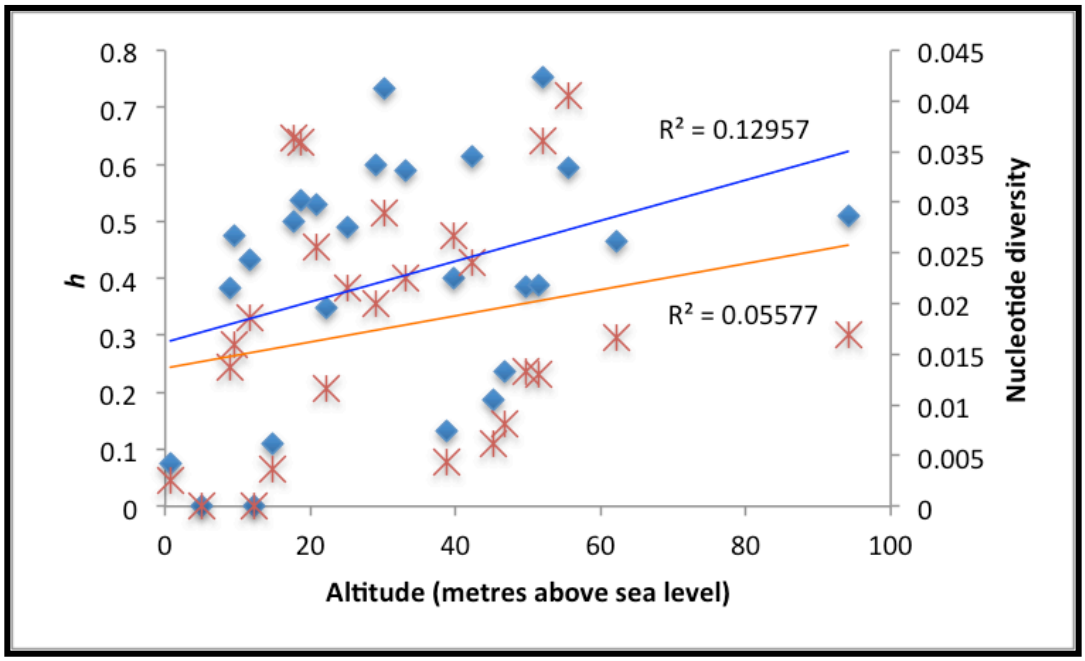


Figure 3.14: Relationship of mtDNA diversity (D-loop) and altitude (elevation above sea level (metres)) in roach. Blue diamonds = haplotypic diversity; red stars = nucleotide diversity.

Regardless of the potential causative factors at play, the actual level of divergence between upstream and downstream populations within rivers was not found to be statistically significant (as opposed to some within river-system divergences), suggesting that mtDNA is limited in its scope to determine fine-scale differentiation within rivers themselves, at least at the sampling resolution employed herein. Therefore this subject shall be revisited in the next chapter by employing genetic markers more attuned for scales of high geographic resolution.

Generally, however, the progression that has been made in expediting nucleotide sequencing at decreasing costs, indicates that soon whole mitochondrial genomes will be made available for comparison between individuals, demes and species on a large scale (Carr and Marshall 2008), and such endeavors have begun (e.g. in killer whales *Orcinus orca* (Morin et al. 2010)). The future of population screening based upon mtDNA variation will inevitably go down this holistic route and will enable a more robust examination of evolutionary and demographic hypotheses.

Appendix A

A.1 Nucleotide Sequences

A.1.1: D-Loop Nucleotide Sequences (5' – 3')

```
D1  GTTAGTACAT ATATATGTAT TATCACCATT CATTTATATT AACCTAAAAG CAAGTACTAA CGTTCAAGAC GTACATAAAG CAAATTGTTA AACTCAGAA ATATTTTATT TTAACTTAAG
D2  .......... .......... .......... .......... .......... .......... .......... .......... .......... ......... .......... .........G.
D3  .......... .......... .......... .......... .......... .......... .......... .......... .......... ......... .......... ..........
D4  .......... .......... .......... .......... .......... .......... .......... .......... .......... ......... .......... ..........
D5  .......... .......... .......... .......... .......... .......... .......... .......... .......... ......... .......... ..........
D6  .......... .......... .......... .......... .......... .......... .......... .........A .......... ......... .......... ..........
D7  .......... .......... .......... .......... .......... .......... .......... .......... .......... ......... .......... ..........
D8  .......... .......... .......... .......... .......... .......... .......... .......... .......... ......... .......... ..........
D9  .......... .......... .......... .......... .......... .......... .......... .......... .......... ......... .......... ..........
D10 .......... .......... .......... .......... .......... .......... .......... .......... .......... ......... .......... ..........
D11 .......... .......... .......... .......... .......... .......... .......... .......... .......... ......... .......... ..........
D12 .......... .......... .......... .......... .......... .......... .......... .......... .......... ......... .......... ..........
D13 .......... .......... .......... .......... .......... ..G....... .......... .......... .......... ......... .......... ..........
D14 .......... .......... .......... .......... .......... .......... .......... .......... .......... ......... .......... ..........
D15 .......... .......... .......... .......... .......... .......... .......... .......... .......... ......... .......... .........G.
D16 .......... .......... .......... .......... .......... .......... .......... .......... .......... ......... .......... ..........
D17 .......... .......... .......... .......... .......... .......... .......... .......... .......... ......... .......... ..........
D18 .......... .......... .......... .......... .......... .......... .......... .......... .......... ......... .......... ..........
D19 .......... .......... .......... .......C.. .......... .......... .......... .......... .......... ......A.. .......... ..........
D20 .......... .......... .......... .......C.. .......... .......... .......... .......... .......... ......A.. .......... ..........
D21 .......... .......... .......... .......C.. .......... .......... .......... .......... .......... ......A.. .......... ..........
D22 .......... .......... .......... .......C.. .......... .......... .......... .......... .......... ......A.. .......... ..........
D23 .......... .......... .......... .......C.. .......... .......... .......... .......... .......... ......A.. .......... ..........
D24 .......... .......... .......... .......C.. .......... .......... .......... .......... .......... ......A.. .......... ..........
D25 .......... .......... .......... .......C.. .......... .......... .......... .......... .......... ......A.. .......... ..........
D26 .......... .......... .......... .......C.. .......... .......... .......... .......... .......... ......A.. .......... ..........
D27 .......... .......... .......... .......C.. .......... .......... .......... .......... .......... ......A.. .......... ..........
D28 .......... .......... .......... .......C.. .......... .......... .......... .......... .......... ......A.. .......... ..........
D29 .......... .......... .......... .......C.. .......... .......... .......... .......... .......... ......A.. .......... ..........
D30 .......... .......... .......... .......C.. .......... .......... .......... .......... .......... ......A.. .......... ..........
D31 .......... .......... .......... .......... .......... .......... .......... .......... .......... ......... .......... ..........
```

```
D1  GTTAGTACAT ATATATGTAT TATCACCATT CATTTATATT AACCTAAAAG CAAGTACTAA CGTTCAAGAC GTACATAAAG CAAATTGTTA AACTCAGAA ATATTTTATT TTAACTTAAG
D32 .......... .......... .......... .......... .......... .......... .......... .......... .......... ...C..... .......... ..........
D33 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D34 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D35 .......... .......... .......... .......... .......... .......... ...C...... .......... .......... .......... .......... ..........
D36 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D37 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D38 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D39 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D40 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........

D1  AAATAGATAA TTCCCCTAGA TATGGATCTC ACATTTTTCC TCGAAATATA CAACTAAGAT TTAGTTTAAT CATATTAATG TAGTAAGAGA CCACCAACCG GTTCATATAA GGCATATTAT
D2  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D3  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D4  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D5  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D6  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D7  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D8  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D9  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D10 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D11 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D12 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D13 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D14 .......... .......... .......... .A........ .......... .......... .......... .......... .......... .......... .......... ..........
D15 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D16 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D17 .......... .......... .......... .......... .....T.... .......... .......... .......... .......... .......... .......... ..........
D18 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D19 .......... .......... ......C... ....G..... .T......C. .......... .........C .......... .......... .......... .......... ..........
D20 .......... .......... ......C... ....G..... .T......C. .......... .........C .......... .......... .......... .......... ..........
D21 .......... .......... ......C... ....G..... .T......C. .......... .........C .......... .......... .......... .......... ..........
D22 .......... .......... ......C... ....G..... .T......C. .......... .......... .......... .......... .......... .......... ..........
D23 .......... .......... ......C... ....G..... .T......C. .......... .........C T........ .......... .......... .......... ..........
D24 .......... .......... ......C... ....G..... .T......C. .......... .........C .......... .......... .......... .......... ..........
D25 .......... .......... ......C... ....G..... .T......C. .......... .......... .......... .......... .......... .......... ..........
D26 .......... .......... ......C... ....G..... .T......C. .......... .........C .......... .......... .......... .......... ..........
D27 .......... .......... ......C... ....G..... .T......C. .......... .......... .......... .......... .......... .......... ..........
D28 .......... .......... ......C... ....G..... .T......C. .......... .......... .......... .......... .......... .......... ..........
D29 .......... .......... ......C... ....G..... .T......C. .......... .........C .......... .......... .......... .......... ..........
D30 .......... .......... ......C... ....G..... .T......C. .......... .......... .......... .......... .......... .......... ..........
D31 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D32 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D33 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
```

```
D1  AAATAGATAA TTCCCCTAGA TATGGATCTC ACATTTTTCC TCGAAATATA CAACTAAGAT TTAGTTTAAT CATATTAATG TAGTAAGAGA CCACCAACCG GTTCATATAA GGCATATTAT
D34 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D35 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D36 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D37 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D38 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D39 .......... .......... .......... .A........ .......... .......... .......... .......... .......... .......... .......... ..........
D40 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........

D1  TAATGATAGA ATCAGGGACA CAACACTAAG ATACGGTATA TTATGAATTA TTCCTTGTAT CTGGTTCTCC TGTCACGTAC AGACCTGTGA AGAATCCATC CTAATTTATT TTCCTTGCAT
D2  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D3  .......... .......... .......... .......... .......... .......... .......... .......... .....C.... .......... .......... ..........
D4  .......... .......... .......... .......... .......... .......... .......... .......... ....T.C... .......... .......... ..........
D5  .......... .......... .......... .......... .......... .......... .......... .......... .....C.... .......... .......... ..........
D6  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D7  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D8  .......... .......... .......... .......... .......... .......... .......... .......... .....C.... .......... .......... ..........
D9  .......... .......... .......... .......... .......... .......... .......... .......... ......C... .......... .......... ..........
D10 .......... .......... .......... .......... .......... .......... .......... .......... ......C... .......... .......... ..........
D11 .......... .......... .......... .......... .......... .......... .......... .......... .....C.... .......... .......... ..........
D12 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D13 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D14 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D15 .......... .......... .......... .......... .......... .......... .......... .......... .....C.... .......... .......... ..........
D16 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D17 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D18 .......... .......... .......... .......... .......... .......... .......... .......... .....A.... .......... .......... ..........
D19 .......... .......... .......... .......... .......... .......... .......... .......... ....T.C... .......... .......... ..........
D20 .......... .......... .......... .......... .......... .......... .......... .......... ....T.C... .......... .......... ..........
D21 .......... .......... .......... .......... .......... .......... .......... .......... ....T.C... .......... .......... ..........
D22 .......... .......... .......... .......... .......... .......... .......... .......... ....T.C... .......... .......... ..........
D23 .......... .......... .......... .......... .......... .......... .......... .......... ....T.C... .......... .......... ..........
D24 .......... .......... .......... .......... .......... .......... .......... .......... ....T.C... .......... .......... ..........
D25 .......... .......... .......... .......... .......... .......... .......... .......... ....T.C... .......... .......... ..........
D26 .......... .......... .......... .......... .......... .......... .......... .......... ....T.C... .......... .......... ..........
D27 .......... .......... .......... .......... .......... .......... .......... .......... ....T.C... .A........ .......... ..........
D28 .......... .......... .......... .......... .......... .......... .......... .......... ....T.C... .......... .......... ..........
D29 .......... .......... .......... .......... .......... .......... .......... .......... ....T.C... .......... .......... ..........
D30 .......... .......... .......... .......... .......... .......... .......... .......... ....T.C... .......... .......... ..........
D31 .......... .......... .......... .......... .......... .......... .......... .......... ......C... .......... .......... ..........
D32 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D33 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D34 .......... .......... .......... .......... .......... .......... .......... .......... ......C... .......... .......... ..........
D35 .......... .......... .......... .......... .......... .......... .......... .......... ......C... .......... .......... ..........
```

```
D1  TAATGATAGA ATCAGGGACA CAACACTAAG ATACGGTATA TTATGAATTA TTCCTTGTAT CTGGTTCTCC TGTCACGTAC AGACCTGTGA AGAATCCATC CTAATTTATT TTCCTTGCAT
D36 .......... .......... .......... .......... .......... .......... .......... .......... ......A... .......... .......... ..........
D37 .......... .......... .......... .......... .......... .......... .......... .......... ....T.C... .......... .......... ..........
D38 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D39 .......... .......... .......... .......... .......... .......... .......... .......... ......C... .......... .......... ..........
D40 .......... .......... .......... .......... .......... .......... .......... .......... ....T..... .......... .......... ..........

D1  CCGGCTACTG GTGTAATTAC ATACTCCGCA TTACCCCACA TGCCGGGCAT TCTTTTATAT GCATAAGGTT CTTTTTTCTG GTTTCCTTTC ACTTTGCATC TCAGAGTGCA GGCACAATTA
D2  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D3  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D4  .......... .......... .......... .......... .......... .......... .......... .......... .......... .....A.... .......... ..........
D5  .......... .......... .......... .......... .......... .......... .....G.... .......... .......... .......... .......... ..........
D6  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D7  .......... .......... .......... .......... .......... .......... .......... .......... .......... .....A.... .......... ..........
D8  .........A. .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D9  .......... .......... .......... .......... .......... .......... .......... .......... .......... .....A.... .......... ..........
D10 .......... .......... .......... .......... .......... .......... .......... .......... .......... .....A.... .......... ..........
D11 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D12 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... C......... ..........
D13 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D14 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D15 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D16 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D17 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D18 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D19 .......... .......... .......... .......... .......... .......... .......... .......... .......... .....A.... .......... .......C..
D20 .......... .......... .......... .......... .......... .......... .......... .......... .......... .....A.... .......... ..........
D21 .......... .......... .......... .......... .......... .......... .......... .......... .......... .....A.... .......... .......C..
D22 .......... .......... .......... .......... .......... .......... .......... .......... .......... .....A.... .......... ..........
D23 .......... .......... .......... .......... .......... .......... .......... .......... .......... .....A.... .......... ...G...C..
D24 .........A. .......... .......... .......... .......... .......... .......... .......... .......... .....A.... .......... ..........
D25 .......... .......... .......... .......... .......... .......... .......... .......... .......... .....A.... .......... ...G...C..
D26 .......... .......... .......... .......... .......... .......... .......... .......... .......... .....A.... .......... ...G...C..
D27 .......... .......... .......... .......... .......... .......... .......... .......... .......... .....A.... .......... .......C..
D28 .......... .......... .......... .......... .......... .......... .....G.... .......... .......... .....A.... .......... .......C..
D29 .......... .......... .......... .......... .......... .......... .....G.... .......... .......... .....A.... .......... ..........
D30 .......... .......... .......... .......... .......... .......... .......... .......... .......... .....A.... .......... .......C..
D31 .......A.. .......... .......... .......... .......... .......... .....G.... .......... .......... .....A.... .......... ..........
D32 .......... .......... .......... .......... .......... .......... .......... .......... .......... .....A.... .......... ..........
D33 .......... .......... .......... .......... .......... .......... .......... .......... .......... .....A.... .......... ..........
D34 .......... .......... .......... .......... .......... .......... .....G.... .......... .......... .......... .......... ..........
D35 .......... .......... .......... .......... .......... .......... .......... .......... .......... .....A.... .......... ..........
D36 .......... .......... .......... .......... .......... .......... .......... .......... .......... .....A.... .......... ..........
D37 .......... .......... .......... .......... .......... .......... .......... .......... .......... .....A.... .......... ..........
```

```
D1  CCGGCTACTG GTGTAATTAC ATACTCCGCA TTACCCCACA TGCCGGGCAT TCTTTTATAT GCATAAGGTT CTTTTTTCTG GTTTCCTTTC ACTTTGCATC TCAGAGTGCA GGCACAATTA
D38 .......... .......... .......... .......... .......... .......... .....G.... .......... .......... .....A.... .......... ..........
D39 .......... .......... .......... .......... .......... .......... .......... .......... .......... .....A.... .......... ..........
D40 .......... .......... .......... .......... .......... .......... .......... .......... .......... .....A.... .......... ..........

D1  ATATATCAA GGTAGTATAT TTCCTTGCAA GAGTTAAAGT TGGTTCATTA TTAAAAGACA TAACTTAAGA ATTACATATT ATTTTATCAA GTGCATAATA CATTCATCTC TTCTTCCAAC
D2  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D3  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D4  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D5  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D6  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D7  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D8  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D9  .......... ...G...... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D10 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D11 .......... .......... .......... .......... C......... .......... .......... .......... .......... .......... .......... ..........
D12 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D13 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D14 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D15 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D16 .......... .......... .......... .......... C......... .......... .......... .......... .......... .......... .......... ..........
D17 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D18 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D19 .......... ...G...C.. .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D20 .......... ...G...C.. .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D21 .......... ...G...C.. .......... .......... .......... ..G....... .......... .......... .......... .......... .......... ..........
D22 .......... ...G...C.. .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D23 .......... ...G...C.. .......... .......... .......... ..G....... .......... .......... .......... .......... .......... ..........
D24 .......... ...G...C.. .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D25 .......... ...G...C.. .......... .......... .......... ..G....... .......... .......... .......... .......... .......... ..........
D26 .......... ...G...C.. .......... .......... .......... ..G....... .......... .......... .......... .......... .......... ..........
D27 .......... ...G...C.. .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D28 .......... ...G...C.. .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D29 .......... ...G...C.. .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D30 .......... ...G...C.. .......... .......... .......... ..G....... .......... .......... .......... .......... .......... ..........
D31 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D32 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D33 .......... ...G...... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D34 .......... ...G...... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D35 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D36 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D37 .......... ...G...... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D38 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
D39 .......... ...G...... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
```

```
D1  ATATATCAA GGTAGTATAT TTCCTTGCAA GAGTTAAAGT TGGTTCATTA TTAAAAGACA TAACTTAAGA ATTACATATT ATTTTATCAA GTGCATAATA CATTCATCTC TTCTTCCAAC
D40 ......... ......... ......... ......... ......... ......... ......... ......... ......... ......... ......... .........

D1  TTACCCCTAT ATATATGCCC CCCCTTTTGG CTTTTG
D2  .......... .......... .......... ......
D3  .......... .......... .......... ......
D4  .......... .......... .......... ......
D5  .......... .......... .......... ......
D6  .......... .......... .......... ......
D7  .......... .......... .......... ......
D8  .......... .......... .......... ......
D9  .......... .......... .......... ......
D10 .......... .......... .......... ......
D11 .......... .......... .......... ......
D12 .......... .......... .......... ......
D13 .......... .......... .......... ......
D14 .......... .......... .......... ......
D15 .......... .......... .......... ......
D16 .......... .......... .......... ......
D17 .......... .......... .......... ......
D18 .......... .......... .......... ......
D19 .......... .......... .......... ......
D20 .......... .......... .......... ......
D21 .......... .......... .......... ......
D22 .......... .......... .......... ......
D23 .......... .......... .......... ......
D24 .......... .......... .......... ......
D25 .......... .......... .......... ......
D26 .......... .......... .......... ......
D27 .......... .......... .......... ......
D28 .......... .......... .......... ......
D29 .......... .......... .......... ......
D30 .......... .......... .......... ......
D31 .......... .......... .......... ......
D32 .......... .......... .......... ......
D33 .......... .......... .......... ......
D34 .......... .......... .......... ......
D35 .......... .......... .......... ......
D36 .......... .......... .......... ......
D37 .......... .......... .......... ......
D38 .......... .......... .......... ......
D39 .......... .......... .......... ......
D40 .......... .......... .......... ......
```

A.1.2: Cytochrome B Nucleotide Sequences (5' – 3')

```
NEW    ATG GCA AGC CTA CGA AAA ACC CAT CCA CTA ATA AAA ATC GCT AAT GAC GCG CTA GTC GAC CTT CCG ACA CCA TCT AAC ATC TCA GCA CTA TGA AAC TTC
UKSS   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
UK     ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
KenSo  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
UKTh   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
UKTe   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

NEW    GGG TCC CTG CTA GGG TTA TGT TTA ATT ACC CAA ATC CTG ACA GGA CTA TTC TTA GCT ATA CAC TAT ACC TCT GAC ATC TCA ACC GCG TTT TCA TCG GTG
UKSS   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
UK     ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
KenSo  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
UKTh   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
UKTe   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

NEW    ACC CAC ATC TGC CGA GAC GTC AAC TAC GGC TGA CTT ATC CGA AAC CTA CAT GCT AAT GGA GCA TCC TTC TTC TTC ATC TGT CTT TAT ATA CAT ATC GCA
UKSS   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..G ... ... ... ... ... ... ... ... ... ... ...
UK     ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
KenSo  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
UKTh   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
UKTe   ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

NEW    CGA GGC CTA TAT TAC GGG TCA TAC CTT TAT AAG GAA ACC TGA AAC ATT GGT GTG GTT CTA TTC CTC CTG GTT ATA ATG ACA GCC TTC GTT GGC TAC GTA
UKSS   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
UK     ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... A.. ... ... ... ... ... ... ... ... ... ... ...
KenSo  ... ... ... ... ... ... ... ... ... ... ... ... G.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
UKTh   ... ... ... ... ... ... ... ... ... ... ..A ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
UKTe   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

NEW    CTA CCA TGA GGA CAA ATA TCA TTC TGA GG
UKSS   ... ... ... ... ... ... ... ... .. 
UK     ... ... ... ... ... ... ... ... .. 
KenSo  ... ... ... ... ... ... ... ... .. 
UKTh   ... ... ... ... ... ... ... ... .. 
UKTe   ... ... ... ... ... ... ... ... .. 
```

A.2 European Roach Data

A.2.1: Table of sampling information for European roach samples

Sampling Location	Code	Co-ordinates		Sampling Date
		Latitude	Longitude	
Lake Kottegat, Denmark	DENA	U	U	2008
Unknown Lake, Denmark	DENB	U	U	2008
Sude de Lyon, River Rhône, France	RH	45.590000	4.775556	2009
River L'Ancre, Saone, France	SA	50.074444	2.630278	U
Lake Höytiäinen, Finland	FIN	62.666667	29.500000	2008
River Isar, Danube, Germany	ID	48.848610	12.985560	20/08/2004
Altmühl, Danube, Germany	AD	49.318056	10.707778	06/09/2004
Amper Canal, Danube, Germany	ACD	48.527500	11.880556	20/07/2006
River Saale, Main, Germany	SM	U	U	19/09/2005
River Wern, Main, Germany	WM	50.059444	10.054444	14/09/2005
River Main, Germany	M	50.073611	11.041944	17/11/2004
River Elbe	ELB	U	U	U
Eger Stream, Elbe	EES	U	U	22/09/2007
Sächsische Saale nr Förbau, Elbe, Germany	ESSF	U	U	26/09/2007
Selb Creek, Elbe, Germany	ESC	U	U	22/09/2007
Selbitz, Elbe, Germany	ES	U	U	23/10/2007
Sächsische Saale nr Joditz, Elbe, Germany	ESSJ	U	U	23/10/2007
Caspian Sea, off Kazhakstan	CAS	38.245015	46.970219	U

U = unknown.

A.2.2: Table of D-loop haplotype counts for all European sampling locations (see A.3.1 for location code details).

Pop	Haplotype																		
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16	D17	D18	D19	D20
DENA	3			2																
DENB			1	2																
RH	7						22						1			3				
SA	5	1		1						1										
FIN	4			1																
ID	8		4	1					3											
AD	4						1			1										
ACD	1																			
SM	12		3	2					4											
WM	3			1						1										
M	5		3						1											
ELB				1																
EES			2						1	1										
ESSF	1			1																
ESC	1																			
ES	2		1																	
ESSJ	3																			
CAS																			2	10

A.2.2: Continued.

Pop	Haplotype																			
	D21	D22	D23	D24	D25	D26	D27	D28	D29	D30	D31	D32	D33	D34	D35	D36	D37	D38	D39	D40
DENA																				
DENB																				
RH												1								
SA																				
FIN											1									
ID													1	1	1	1				
AD																	1			
ACD																				
SM																			1	
WM													1				1	1		
M													1		1					
ELB																				1
EES																				
ESSF																				
ESC																				
ES																				
ESSJ																				
CAS	2	2	1	2	1	1	1	1	3	1										

A.2.3: Table of D-loop diversity metrics for all European sampling locations.

Sampling Location	Hap diversity		Nuc diversity	
	h	**s.d.**	**π**	**s.d.**
Lake Kottegat, Denmark	0.600	0.175	0.060	0.018
Unknown Lake, Denmark	0.667	0.314	0.089	0.042
Sude de Lyon, River Rhône, France	0.545	0.085	0.025	0.005
River L'Ancre, Saone, France	0.643	0.184	0.045	0.016
Lake Höytiäinen, Finland	0.600	0.215	0.069	0.025
River Isar, Danube, Germany	0.805	0.07	0.067	0.009
Altmühl, Danube, Germany	0.714	0.181	0.054	0.019
Amper Canal, Danube, Germany	NC*	NC	NC	NC
River Saale, Main, Germany	0.671	0.091	0.060	0.010
River Wern, Main, Germany	0.893	0.111	0.073	0.011
River Main, Germany	0.764	0.107	0.057	0.014
River Elbe	1.000	0.500	0.033	0.017
Eger Stream, Elbe	0.833	0.222	0.083	0.024
Sächsische Saale nr Förbau, Elbe, Germany	1.000	0.500	0.100	0.050
Selb Creek, Elbe, Germany	NC	NC	NC	NC
Selbitz, Elbe, Germany	0.667	0.314	0.022	0.011
Sächsische Saale nr Joditz, Elbe, Germany	0	0	0	0
Caspian Sea, off Kazhakstan	0.852	0.059	0.063	0.009
Σ	**11.254**	**3.128**	**0.901**	**0.278**
μ	**0.703**	**0.196**	**0.056**	**0.017**

*NC = Not calculated.

A.3 Roach Clade P-Distances

A.3.1: Table of p-distances between all discovered D-loop haplotypes in all European and UK common roach samples.

	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16	D17	D18	D31	D32	D33	D34	D35	D36	D37	D38	D39
D1																											
D2	0.002																										
D3	0.002	0.003																									
D4	0.005	0.006	0.006																								
D5	0.003	0.005	0.002	0.008																							
D6	0.002	0.003	0.003	0.006	0.005																						
D7	0.002	0.003	0.003	0.003	0.005	0.003																					
D8	0.003	0.005	0.002	0.008	0.003	0.005	0.005																				
D9	0.005	0.006	0.006	0.003	0.008	0.006	0.003	0.008																			
D10	0.003	0.005	0.005	0.002	0.006	0.005	0.002	0.006	0.002																		
D11	0.003	0.005	0.002	0.008	0.003	0.005	0.005	0.003	0.008	0.006																	
D12	0.002	0.003	0.003	0.006	0.005	0.003	0.003	0.005	0.006	0.005	0.005																
D13	0.002	0.003	0.003	0.006	0.005	0.003	0.003	0.005	0.006	0.005	0.005	0.003															
D14	0.002	0.003	0.003	0.006	0.005	0.003	0.003	0.005	0.006	0.005	0.005	0.003	0.003														
D15	0.003	0.002	0.002	0.008	0.003	0.005	0.005	0.003	0.008	0.006	0.003	0.005	0.005	0.005													
D16	0.002	0.003	0.003	0.006	0.005	0.003	0.003	0.005	0.006	0.005	0.002	0.003	0.003	0.003	0.005												
D17	0.002	0.003	0.003	0.006	0.005	0.003	0.003	0.005	0.006	0.005	0.005	0.003	0.003	0.003	0.005	0.003											
D18	0.002	0.003	0.002	0.006	0.003	0.003	0.003	0.003	0.006	0.005	0.003	0.003	0.003	0.003	0.003	0.003	0.003										
D31	0.006	0.008	0.008	0.005	0.006	0.008	0.005	0.009	0.005	0.003	0.009	0.008	0.008	0.008	0.009	0.008	0.008	0.008									
D32	0.003	0.005	0.005	0.005	0.006	0.005	0.002	0.006	0.005	0.003	0.006	0.005	0.005	0.005	0.006	0.005	0.005	0.005	0.006								
D33	0.003	0.005	0.005	0.005	0.006	0.005	0.002	0.006	0.002	0.003	0.006	0.005	0.005	0.005	0.006	0.005	0.005	0.005	0.006	0.003							
D34	0.005	0.006	0.006	0.006	0.005	0.006	0.006	0.008	0.003	0.005	0.008	0.006	0.006	0.006	0.008	0.006	0.006	0.006	0.005	0.008	0.005						
D35	0.005	0.006	0.006	0.003	0.008	0.006	0.003	0.008	0.003	0.002	0.008	0.006	0.006	0.006	0.008	0.006	0.006	0.006	0.005	0.005	0.005	0.006					
D36	0.003	0.005	0.005	0.003	0.006	0.005	0.002	0.006	0.003	0.002	0.006	0.005	0.005	0.005	0.006	0.005	0.005	0.005	0.005	0.003	0.003	0.006	0.003				
D37	0.006	0.008	0.008	0.002	0.009	0.008	0.005	0.009	0.002	0.003	0.009	0.008	0.008	0.008	0.009	0.008	0.008	0.008	0.006	0.006	0.003	0.005	0.005	0.005			
D38	0.003	0.005	0.005	0.005	0.003	0.005	0.002	0.006	0.005	0.003	0.006	0.005	0.005	0.005	0.006	0.005	0.005	0.005	0.003	0.003	0.003	0.005	0.005	0.003	0.006		
D39	0.006	0.008	0.008	0.005	0.009	0.008	0.005	0.009	0.002	0.003	0.009	0.008	0.008	0.005	0.009	0.008	0.008	0.008	0.006	0.006	0.003	0.005	0.005	0.005	0.003	0.006	
D40	0.003	0.005	0.005	0.002	0.006	0.005	0.002	0.006	0.005	0.003	0.006	0.005	0.005	0.005	0.006	0.005	0.005	0.005	0.006	0.003	0.003	0.008	0.005	0.003	0.003	0.003	0.006

Chapter Four – Within-Population Microsatellite DNA Diversity and Among-Population Differentiation of Roach Within and Between Two Historically Connected Rivers in the Southeast of England

"What should I speake of the fat and sweet salmons, dailie taken in this streame, and that in such plenty (after the time of the smelt be past) as no river in Europa is able to exceed it. What store also of barbells, trouts, cheuins, pearches, smelts, breams, roches, daces, gudgings, flounders, shrimps, etc. are commonlie to be had herein…this famous riuer complaineth commonlie of no want, but the more it looseth at one time, the more it yeeldeth at another" **–** ***Clergyman and writer William Harrison on the bountiful fishery of the Thames – A Description of England (1577), quoted in Ackroyd (2009).***

4.1 General Introduction

Despite possessing relatively limited levels of nucleotide diversity within the mitochondrial genome, the roach exhibits detectable levels of haplotype diversity within and among a number of drainages within its UK distribution. However, a single line of evidence does not provide a robust conclusion about the distribution of overall levels of genetic variation within the United Kingdom, upon which the future evolutionary potential of the roach is critically dependent. A second line of evidence can be drawn upon, arising from a survey of nuclear variation, to not only bolster hypothesized MUs, but to also investigate fine-scale microevolutionary and demographic processes that affect spatially proximate populations. It is results from these kinds of genetic surveys upon which modern population management methods are becoming increasingly reliant.

The following introductory section will introduce the basic rationale for applying population genetic and molecular ecological techniques to the management and

conservation of threatened and economically valuable organisms, with an emphasis upon freshwater fisheries. The overview will examine in some detail how genetic analyses may help determine broad-scale patterns of colonization and re-colonization (e.g. in metapopulation systems with source-sink dynamics), midscale patterning of inter-population gene flow (often allied to the physical environment, e.g. presence or absence of barriers to physical dispersal), and even in the elucidation of neighborhood sizes and breeding units. Without the use of molecular genetic markers, many of these highly informative ecological factors may not be uncovered; or the elucidation of such phenomena through traditional experimental means would greatly increase the financial overheads of management schemes. This epitome shall also briefly touch upon theoretical and methodological issues associated with investigating the fine-scale population genetics of a freshwater fish such as the roach. Finally, an overview of the current literature of roach population genetics is set out, with a focus on levels of structuring heretofore discovered.

4.1.1 Population Structure and Estimates of Inter-Population Migration

The delineation of population structure has myriad applications to pure and applied population biology, and to the conservation and management of biological populations, including the determination of MUs and studying the effects of habitat fragmentation and differential selection regimes (Hauser & Carvalho 2008). The main problems with estimates of population structure inferred from data of population allele frequencies (i.e. the statistical discrimination of allele frequencies) within species, that are outwardly difficult to delineate into phenotypically or geographically distinct populations, are allied to issues of statistical power: low power may not detect population structure when it *is* present, whereas too high a discriminative power may infer structure when it *is not* present or when it is negligible (see Palsbøll et al. 2010). Poor sampling schemes may also overestimate estimates of population structure (Schwartz & McKelvey 2009). Alternatively, it is possible to both define and measure population structure as resulting from a reduction - or cessation - of inter-population gene flow. As Palsbøll and colleagues (2010) point out, dispersal may also be inferred from population genetic data. Differentiating between recent dispersal and long-term migration (gene flow) provide two useful and complementary metrics of inter-population transfer (one measures the ability or efficacy to disperse to another location, whilst the other measures the ability or efficacy to disperse *and* reproduce over longer ecologically and evolutionarily significant

timeframes), whose measurement and application have differing consequences for management and conservation concerns. Confusingly, however, the term 'dispersal' is used as a blanket term in the literature to cover most aspects of geographical connectivity among populations, including those instances in which fecund adults assimilate their gene frequencies into the recipient gene pool via *in situ* mating with resident individuals; and first-generation migrants, who may or may not successfully interbreed with conspecifics in a recipient deme. Further, gene flow may also be mediated by passive germ cell and larval 'dispersal', in addition to the translocation of fecund adult individuals (e.g. Peterka et al. 2004).

Once population structure has been inferred, one may then embark upon estimating both the extent and rate of effective dispersal, as well as identifying recent immigrants (Davies et al. 1999), that is an important determinant of much of the spatial distribution of demes, individuals and genes (Broquet and Petit 2009). The oldest method of determining migration rates from genetic data derives from Wright's *F*-statistics, in which *Nm* (the local immigration rate) is calculated (Wang and Whitlock 2003) (see Chapter Two). The current literature favours the 'population-based' methods of assignment (Paetkau et al. 1995; Rannala & Mountain 1997; Wilson & Rannala 2003; Faubet & Gaggiotti 2008; and Broquet and Petit 2009) and coalescent and likelihood-based (Beerli & Felsenstein 2001; Nielsen & Wakeley 2001) inferences of population structure. Once the extent to which populations are sub-divided is determined, and the direction of gene flow and the degree to which populations exchange genetic material is inferred, how is such information assimilated within a management context?

4.1.2 The Incorporation of High Resolution Population Genetic Data as a Valuable Management Tool

The issue of the concept of 'stocks' or 'populations' in population biology has been dealt with in Chapter Two. Following Ihssen et al.'s formulation of a 'stock' (1981), a population should be interpreted as an interbreeding aggregate of individuals that possesses some cohesive genetic and spatial tractability. Therefore, this group of individuals may still exchange effective migrants at a rate up to some level at which both the spatial and genetic cohesiveness of the exchanging populations breaks down into homogeneity (following Palsbøll et al.'s (2010) second definition of population structuring). The significance of whether species exist, as a single homogeneous,

panmictic collection of individuals or as a number of sub-divided, panmictic populations, is significant for both management and conservation concerns.

Single large populations are less likely to go extinct in the face of unpredictable, stochastic environmental and demographic factors, whereas the risk for equally large but fragmented population is considerably greater, a theoretical proposition (MacArthur & Wilson 1967) supported empirically across a range of taxa (Reed 2004), including freshwater fish (e.g. Fagan et al. 2002, but see below for a detailed exposition). In small or fragmented populations, the related genetic processes of genetic drift and population inbreeding are greatly increased in effectiveness such that together these processes work in synergism so that they become detrimental to a population's chance of continued, temporal persistence (Franklin 1980; Lande 1994; and O'Grady et al. 2006). However, the risk of inbreeding in large panmictic populations is negligible in the face of recurrent mutation or migration, and localized if it does occur[25]. Sampling effects in small populations may even fix severely detrimental alleles through random fixation despite the presence of directional selection against any negative phenotypic effect (Lande 1988).

Low levels of genetic diversity within a population substantially increase the risk of its extinction (Lande 1988). Low census numbers contribute to extinction risk through an increased susceptibility to stochastic environmental catastrophes, and because of a reduction in the number of successful matings[26]. Small isolated populations also are more susceptible to attack by opportunistic pathogens and parasites: reduction in the heterozygosity of the MHC complex severely restricts the number of antigen proteins available for expression on the cell surfaces of white blood cells, which in turn increases the likelihood that some bacterium or virus will overcome this line of defense. Whilst it is true that infectious diseases spread more efficiently in large, dense populations, natural selection also is more effective in such populations and so will often result in long-term population persistence. Small populations on the other hand, even if alleles for resistance are present, may be susceptible to the strong influence of genetic drift increasing extinction risk even in face of selection for resistance (the selection coefficient needs to be greater than the chance of a random allele reaching fixation, i.e. $s >> 1/2N$ in a diploid population). If populations become so small as to be susceptible to the effects of

[25] The time for the fixation of any allele is much greater (it is a product of the inverse of two times the population size, for nuclear alleles in diploid genomes)

[26] Sexually mature adults may never locate one another during the breeding season(s). Also, reduced census numbers increase the possibility of a reduction in fertility or offspring viability due to inbreeding depression.

demographic, environmental and genetic stochasticity, then they are said to enter an extinction vortex (Gilpin & Soulé 1986; Fagan & Holmes 2006) in which extinction is inevitable unless recuperative action is taken.

In small populations the potential for the rejuvenating effect of inter-population migration becomes very important (Hoehn et al. 2007). In source-sink metapopulations recurrent natural recolonization often compensates for localized extinctions of sub-populations. The anthropogenic restitution of gene flow is termed genetic restoration and this is implemented when it has become clear that no such natural compensatory dispersal is possible. The presence of migration from one population to another may allow populations that are undergoing environmental stresses to persist (Kokko & López-Sepulcre 2006). On the other hand, if fragmented populations experience little migration, but still contain high census sizes in a heterogeneous environmental landscape, then such populations may accumulate and retain more genetic diversity, a repository of which provides a benefit for the species in future times of environmental uncertainty. In such cases, recent migration, either natural (e.g. due to flooding between previously isolated streams or pools) or anthropogenic in nature, may erode adaptive differences among populations. This may have a detrimental effect on not only the future success of the species as a whole, but on the continued success of the local population, as adaptive gene complexes are broken up through the interbreeding of indigenous with exogenous individuals (genetic introgression). In instances when such translocations or mediated migrations are anthropogenic in nature, this is called, with some justification, genetic pollution.

For the above reasons (but see Chapter Two), the primary justification of population management is to preserve and protect the diversity of biological populations, in addition to securing ecosystem services for human populations and preserving biodiversity for posterity from a philosophical and moral standpoint. This manifesto invariably includes the preservation of genetic diversity, often hidden and not expressed phenotypically (Le Rouzic & Carlborg 2008), both within and among populations. Therefore, genetic methods are necessary to uncover this variation, quantify it, qualify it in terms of spatial and temporal distribution, and to infer from it processes that are either experimentally intractable or financially prohibitive to employ wholesale mark-recapture techniques (Lande & Barrowclough 1987; DeYoung & Honeycutt 2005). Population genetic information may also inform upon the potential for the future success and longevity of a particular population or species (e.g. in postulating 'minimal viable populations', or

MVPs, based upon the probability of a population surviving into the future (Gilpin & Soulé 1986)).

It is arguable that the most important variable in the current management of populations, which has a strong potential influence upon all of the evolutionary and ecological pitfalls and processes that may befall a species, is migration and gene flow: the long-term dispersal and short-term movements of individuals (Trakhtenbrot et al. 2005). Most wildlife management institutions deal with timescales that incorporate generations of individuals, and the greatest contributor to genetic variation within populations, and the level of structure among populations, at such high-resolution evolutionary scales is effective migration.

4.1.3 Threats to Freshwater Fish: Implications for Management

Freshwater fish, along with other organisms found within riverine landscapes, inhabit a geometric physical environment that is somewhat at odds with those of terrestrial and marine organisms (Fagan 2002; Fagan et al. 2009). Due to the single-dimension[27] that characterizes the dendritic geometry of river systems, thus imposing directional limits upon routes and avenues for inter-demic dispersal, the risks posed to population longevity by the destructive effects of habitat fragmentation are greatly increased (Fagan 2002). Riverine habitats tend to exhibit a natural heterogeneity that influences biodiversity patterning (Fagan 2002; Fagan et al. 2009), such that destruction of habitats will exacerbate the 'natural patchiness' of the already heterogeneous dendritic environment. Habitat destruction, resulting in the loss of suitable communal feeding areas and, more fundamentally, optimal spawning sites, is perhaps the greatest long-term threat to the persistence of a fishery within a river system. The loss of both feeding areas and spawning sites draws the affected species into direct competition with other species that may not have been natural competitors previously. A downturn in breeding success will inevitably lead to lower rates of recruitment, thus driving down census numbers and increasing the susceptibility of these smaller populations to the demographic and genetic problems outlined previously. Additionally, where closely related species inhabit similar habitats and possess similar spawning preferences, the potential deleterious effect of hybridization increases (Rhymer & Simberloff 1996).

[27] The single dimensionality of river systems is, of course, an oversimplification of reality, but is a useful descriptor when couching genetic diversity of riverine organisms, particularly large taxa, in terms of their limited dispersal options.

The proximity of habitat patches, and by corollary of self-sustaining populations of breeding individuals, is an important contributing factor to the occurrence of dispersal between populations (Urban and Keitt 2001**).** In metapopulations, in which some populations are considerably more important sources of migrants than others, the cessation of migrants may have a potentially drastic impact upon population sustainability (Fagan 2002). This is particularly the case in river catchments that periodically dry out and in which isolated tributaries are repopulated only when the connective downstream waters are re-flooded after times of severe drought. However, in demographically well-populated species, little is known of the source of migrants over the long term or over a period encompassing a few generations, or whether dispersal plays an important role. If long-distance dispersal (LDD) contributes to a large part of the gene pool of local populations (as seems to be the case in many instances (Trakhtenbrot et al. 2005)), then the negative effects of significant barriers to dispersal are potentially extremely injurious. Additionally, stochastic levels of gene flow may in some periods be catastrophically low and highly unpredictable, such that local extinctions may occur without any direct contribution from the immediate habitat quality or quantity. The maintenance of an *in situ* repository of population and genetic restitution, that would help local fisheries stocks to recover after such events, is of prime importance. Traditionally, the response has been to re-stock from elsewhere, but as discussed later this may be a waste of resources that could be put to better, more-informed use.

4.1.4 River Networks and Genetic Structuring

By their very nature, river systems limit the potential exchange of reproductive individuals among populations via facilitated dispersal along a unidimensional transect, and this should theoretically translate into potentially greater levels of genetic structuring between distally-located populations than that found in traditional two-dimensional environments such as the marine realm, where there may exist fewer barriers to migration and larger effective population sizes (Gyllensten 1985; Ward et al. 1994). Whilst this simplistic assumption regarding marine populations is less universal than previously thought, levels of genetic differentiation within freshwater fish species are in general much higher and tend to be significant among river catchments and among tributaries across a range of taxa with varying ecologies and life history characteristics (e.g. Hänfling & Brandl 1998; Triantafyllidis et al. 2002; Huey et al. 2010; and Sharma & Hughes 2011), and especially

in rare species (e.g. the endangered cyprinids, *Chondrostoma lusitanicum* (Mesquita et al. 2001) and *Anaecypris hispanica* (Salgueiro et al. 2003)). For many species this is the case, but in some freshwater fish that do have the ability to move between catchments[28] estimates of genetic diversity among drainages are on the scale that one may consider to be low (e.g. in the Australian Gobiid *Hypseleotris compressa* estimates of inter-catchment F_{ST} values were low for both mitochondrial and allozyme data (McGlashan & Hughes 2001)).

4.1.5 Population Genetic Studies of the Common Roach

4.1.5.1 Allozymes and RAPD Variation in Common Roach

The history of genetic studies of wild populations of roach has been an ongoing concern for the last two decades, with a focus upon populations within the great rivers of central Europe. The first use of nuclear-inherited genetic markers was in a 1976 paper that was able to distinguish the systematic inter-relationships among roach and its confamilial species, rudd and bream, by employing allozymes to the problem of taxonomy (Brassington & Ferguson 1976). However, the first population-level application of allozyme loci did not occur until the mid-1980s when Evlanov (1986) was able to apportion roach individuals within a single lake to two separate stocks. As can be seen from the published studies listed in Table 4.1, the application of molecular ecology techniques and population genetic analytical methods to investigate the distribution of genetic diversity within and among wild populations of roach did not occur until the 1990s.

Bouvet and colleagues (1991) carried out the first study of population genetics in roach to directly test the influence of physical habitat upon population structure, as quantified by 10 polymorphic allozyme loci. The eight sampling locations took in sites both above and below a large hydroelectric dam in the upper reaches of the River Rhône, in southern France, in addition to a population of roach in Lake Geneva (the source of the Rhône). Although levels of genetic diversity between populations was low compared to that within populations - 90% of the variation in allele frequencies was allocated to within populations (levels of polymorphism range from 21 – 31% and gene diversity ranging from 0.097 –

[28] Some stenohaline fish, such as the roach, may be able to survive estuarine conditions and populate nearby drainages in which it shares a coastline.

0.124) – the data identified three population groupings based upon Nei's (1973) genetic distance, D: three Lake Geneva samples and an upstream Rhône population sampled just downstream of the confluence of the Rhône with the Ain (D = 0.048 from the other two groups); a side-arm population located above the confluence of the Ain and the Rhône, but before the city of Lyon; and the lower Rhône populations located downstream of Lyon (D between the latter groupings = 0.021).

Table 4.1: Published studies conducted into the distribution of nuclear genetic diversity in the common roach

Marker	Loci	Location	No. sites	Reference
Allozyme	10	France	12	Bouvet et al. 1991
Allozyme	9	France	8	Bouvet et al. 1995
Allozyme	6	Austria	7	Baranyi et al. 1997
Allozyme	10	France	50	Laroche et al. 1999
Allozyme	13	Germany	32	Wolter 1999
RAPD	18	Germany	7	Wolter et al. 2003
Allozyme	12	Germany	24	Hänfling et al. 2004
Microsatellite	5	Sweden	8	Demandt & Björklund 2007
Microsatellite	5	Sweden	5	Demandt 2010

The stretch of the Rhône that flows to the confluence with the Ain is subject to eight dams (Pattee 1988). It is likely that the divergence of the upper Rhône populations is due in some part to a reduction in migration as would be expected with large scale obstructions within a limited habitat. The side arm population of Trou Louis was divergent from both an upstream and a downstream population (Ferrande and Ford, respectively) located in the same side arm, Lône des Pêcheurs, but was allocated within the lower Rhône group. Interestingly, the study of Bouvet et al. (1991) sampled the divergent side-arm population of Trou Louis at two temporal periods: once during the spawning period (June 1985) and once before the onset of winter (October 1986), during which differing frequencies of alleles were detected. This side arm is a known spawning ground (Bouvet et al. 1985), therefore the temporal genetic variation may reflect the different ecological characteristics of the populations at the two time periods. The Trou Louis population is, according to estimates of D (pooled dataset of the temporal samples), intermediary between the downstream Rhône roach and the upstream Geneva roach, suggestive of some interbreeding between these two groups, but not of panmixia among sampled locations within this large, complex river.

Bouvet and colleagues followed up their initial investigations with a study of roach diversity in the Saone, a tributary of the River Rhône, in an effort to compare and contrast two distinct river segments upstream of where the Saone and Rhône coalesce. These two rivers differ in both slope (at 0.4%, the Rhone is an order of magnitude steeper than the Saone) and fish assemblage (more rheophilic species inhabit the Rhône than the Saone). Utilizing nine polymorphic loci, with a maximum of four alleles per locus, average heterozygosities were found to be lower in the Saone than in the Rhône (two Saone sites were all < 0.08, whereas the six Rhône sites were all > 0.08). Most diversity in the Saone was found to be within populations (89%) as opposed to that found among populations (11%). As in the 1991 study, the sidearm site of Lône des Pêcheurs contributes much of the genetic diversity in the Rhône. A dendrogram based upon Nei's genetic distance of all populations distinguish the Lake Geneva and upper Rhône roach (from 50km downstream) from another group encompassing the sidearm and adjacent populations and a final group distinguishing the two sites sampled within the Saone. The latter group also contains roach from Lake Bourget that is located within the riverine environment of the lower Rhône, not the upper Saone.

Bouvet et al. (1991) ascribe the uniformly high within-population genetic diversity and low inter-population divergence in the Saone to a uniformly lentic, macrophyte-rich habitat, which facilitates the high exchange of migrants within the main river and its local tributary. The genetic concordance of the Saone populations with Lake Bourget is also explained by similar habitat found in the lake. The Rhône, however, possesses a number of marginalized habitats – or ecotones - that are of especially important significance in highly modified river channels in which suitable habitat is patchier. The findings of Bouvet et al. (1991, 1995) of a distinct signal of genetic diversity in an ecotone sidearm of the Rhône is consonant with Bouvet's observation (1992) that it is a probable reservoir of much of the genetic diversity of the adjacent Rhône, although the possible causes of this locally important source of diversity are not distinguished (i.e. behavioral, selective or anthropogenic).

Laroche and colleagues further subjected the roach populations of the Rhône to increased scrutiny in a 1999 study, once again applying allozyme data to uncover any correlation between genetic diversity and habitat heterogeneity and environmental pollution. Overall genetic diversity was high (mean observed heterozygosity ranged from 0.047 – 0.068), with uniform distribution over the three main areas of study: Saone and tributaries; the

Upper Rhône between Lake Geneva and the confluence with the Saone; and the lower Rhône downstream of the city of Lyon, inclusive of a "chemical corridor" in which levels of dissolved pollutants from the city are concentrated. Levels of population structuring, as measured by F_{ST}, was low (F_{ST} = 0.026), although there was more pronounced differentiation between the pooled sites of the upper and lower Rhône, respectively (F_{ST} = 0.042). No differences whatsoever were found between the Saone roach and those located in the lower Rhône (F_{ST} = 0). Within the subsystems, all F_{ST} values were low, except in the upper Rhône (F_{ST} = 0.038). No isolation by distance effect was found. Additionally, no decrease in either polymorphism or heterozygosity was associated with the polluted reaches of the lower Rhône.

Wolter (1998, 1999) also studied the distribution of spatial genetic diversity as described by allozyme frequencies. As part of a cross-species study, roach along with other cyprinids were sampled along a 177km length of the River Spree, within the larger Oder drainage (itself part of the Elbe watershed). In total, 24 populations were sampled along the Spree transect. Additionally, a further 8 locations were sampled within the Elbe catchment to determine whether greater levels of genetic divergence are observed between rivers. Overall, some 988 individuals were sampled. Mean allozyme gene diversity was high across the Rivers Spree and Oder (0.072) (Wolter 1998). However, the results of the 1999 study were indicative of lower among-population divergence than those that came before. Roach populations from the Oder and Spree were indistinguishable on the basis of Nei's unbiased distance (1978). The use of Wright's hierarchical *F*-statistics (Wright 1931, 1938a) corroborated the fact that high levels of intra-population genetic diversity blurred the genetic distinctiveness of non-neighboring populations (15% of genetic variability attributed to differences among subpopulations; 3.9% attributable to differences among rivers; and 2.1% attributable between the Spree and Oder catchments within the Elbe drainage; all resulting variability (77%) present among individuals within populations). Gene diversity was high across all loci (mean value of 0.072), directly comparable to most estimates found in the preceding roach literature (Wolter 1999). Any signal of genetic structuring among allozyme loci in the Elbe is weak. However, the author considered stocking to be of little concern due to its relative unimportance compared to the economically more valuable salmonids, thereby ignoring a potentially homogenizing source of roach movement among catchments and rivers.

Despite high genetic diversities overall, the allozyme analyses *were* able to determine the presence of some structuring within rivers, at least in the Rhône watershed. A smaller

study of Austrian roach of the upper Danube measured genetic variability at allozyme loci both within a small transect of the upper Danube and at two distal sites in the upstream Drau tributary and the eutrophic, subalpine Lake Wallersee (Baranyi et al. 1997). However, unbiased genetic distances (Nei 1978) suggested little differentiation among the Danube and Drau sites (gene diversity range: 0.195 – 0.224, despite some 1000km of river separating the Danube and Drau sites), but some divergence of the lake roach from all others (the Lake population had the lowest gene diversity (0.195)). This study also found exceedingly high levels of heterozygosity within the Danube, compared to the other studies described ($H_o \approx 0.2$), underscoring its potential as an important repository of biodiversity for this species and freshwater fishes in general. The small number of sampled sites, and the restrictive geographic range of those sites, failed to uncover significant divergence amongst connected sub-populations (the Lake Wallersee population has probably drifted to some extent in the absence of migration, as indicated by larger F_{IS} values across all loci).

Hänfling et al. (2004) utilized 12 polymorphic allozyme loci in a study comparing diversity levels of still, floodplain water bodies compared with the main body of water along the middle Elbe, in central Germany. These 22 populations (of N = 25 individuals) were complemented with five individual roach from the Rhine and Danube drainages, respectively. Global F_{ST} of the entire dataset was 0.036 ($p < 0.001$). The majority of the structuring signal was contributed by differences among the drainages of the Rhine, Danube and Elbe, such that when only the 22 Elbe populations were analyzed $F_{ST} = 0.018$. Genetic distance (Nei 1978) was greatest between the Elbe populations and the Rhine and Danube populations, respectively (0.043±0.004 and 0.021±0005, respectively), although the distance between the Danube and Rhine populations was considerably lower (Nei's D = 0.003). Because the dataset for the non-Elbe populations was limited, one would expect strong sampling effects associated with the small sample sizes of both the Rhine and Danube populations, contributing positive bias to estimates of divergence.

The high levels of heterozygosity found at allozyme loci in the previous studies are emblematic of allozyme variation within roach (high levels of heterozygosity were also found amongst roach within southern German lakes (Wagner 1992, in Baranyi et al. 1997)), and in common cyprinids in general (e.g. Hänfling & Brandl 1998). Therefore, finding structuring within linear stretches of rivers among populations that are likely exchanging migrants at a high level is problematic with allozyme loci that have comparatively low levels of polymorphisms compared to other nuclear genetic markers.

Wolter et al. (2003) adopted a new method - the employment of RAPD markers - as a prelude to a more discriminative genetic analysis. Random amplified polymorphic DNA markers are dominant markers that are detected by deploying a number of oligonucleotide primers that randomly anneal to sites scattered about the genome. If a bit of genome is polymorphic for a sequence that the randomly constructed primer recognizes, then that primer serves as a marker, albeit a dominant one (only presence or absence can be noted, i.e. heterozygotes cannot be distinguished from homozygotes). RAPD is a quick way of appraising levels of diversity. Relatedness amongst individuals/populations can be computed from the degree of band sharing (PCR-bands), and thus genetic distances calculated. Wolter successfully used 18 such RAPD loci upon roach sampled at seven sites situated along a 122km tract of the middle Elbe River. However, roach again displayed high levels of within-population genetic diversity, such that the genetic similarity among roach among all sites was 71%. Wolter et al. (2003) concluded the existence of single panmictic populations of roach specifically, and cyprinids in general, in both the Oder and Elbe rivers, with home ranges spanning 120 - 177km. RAPD markers are generally used as preliminary markers to determine the presence of genetic variation, and are statistically compromised compared to co-dominant markers. Therefore RAPD markers, more than allozymes, may consistently miss more subtle levels of genetic structuring at microgeographic levels.

In summary, high levels of genetic diversity in roach, mostly from allozyme surveys, have been found within populations and limited diversity attributed among physically isolated rivers. Most explanations for low levels of population structuring centre upon historically stable, large population sizes and the presence of environmental uniformity within large stretches of lentic riverine habitat. In support of the historical effective population size hypotheses, most early studies were carried out in roach populations of central and southern France and around the Danube drainages of central and western Europe, areas in which Pleistocene refugia have been postulated for freshwater fish. Thus, these roach populations may have existed *in situ* throughout the climatic oscillations of the last 2.6 million years, resulting in very high effective population sizes (see Chapter Three). Where structuring has been observed to occur, it is normally associated with either a significant occlusion to migration (e.g. large hydro-electric dams) or where ecotonic heterogeneity is known to exist, or between physically distinct drainages via the drift of allele frequencies.

Whilst it is difficult to rule out balancing selection at allozyme loci (Altukhov 1991), the low number of alleles at such markers often means that common alleles are shared

amongst populations, which may result in higher than average levels of heterozygosity (Griffiths & Li 1983). Other possibilities for the general low level of structuring within this species ventured by the authors include high levels of panmixia across large river distances, which do not tally with tag-and-marker study of this species (e.g. Baade & Fredrich (1998)). The fact that anthropogenic translocation in this species has been greatly overlooked by many researchers (Wolter aside), despite the roach's status as an angling staple and its history as a bait and food fish, in addition to canal construction among waterways, is somewhat of a puzzle; but even in the UK, records of stocking events are sparse and not recorded for all rivers (Nigel Hewlett, pers comm). Even given the presence of significant levels of intraspecific genetic variation, most previous studies of allozyme variation in roach have found significant differences within and between river systems, albeit of low magnitude.

4.1.5.2 Microsatellite Variation in Common Roach

The development of microsatellite markers in tandem with an increase in the efficacy of genotyping and analytical methods opened up a new avenue of potentially greater variation with which to detect levels of population sub-structuring in the roach and other cyprinids. However, very few such studies have been published in the roach, or in coarse fish in general. To date, only two microsatellite studies of *Rutilus rutilus sensu stricto* have been published (Demandt and Björkland 2007; Demandt 2010), which focus upon closed, lentic populations in Swedish lakes. The first study found a significant reduction in microsatellite diversity at five loci in a population recently restituted with introduced roach compared to wild, control populations that were free of recent reintroductions. Both levels of allelic richness (A_r) and the allelic size range (R) of the microsatellites were significantly greater in the wild, control populations compared to the reintroduced populations. These associations remained significant even after accounting for physical differences amongst the sampled lake populations (such as lake size, lake depth and species diversity). The results are important for management concerns, as despite the large numbers of reintroduced roach, they resulted in the bottlenecking of genetic diversity at the reintroduced sites, probably as a result of restocking from a limited sample of the lake's diversity (a founder effect), and because of differential survival and reproduction amongst the reintroduced animals themselves (a selection effect) (Demandt and Björkland 2007).

In the second study, Demandt (2010) found that, at least for roach within a single, closed basin (Biotest basin at Forsmark, Sweden) used as a receptacle for cooling water from the adjacent nuclear power plant, levels of gene diversity remained stable over the time period 1977 – 2000 (0.73 – 0.80), although F_{ST} differed significantly in this time (Mantel test of genetic distance with time as measured at five points in time (1977, 1982, 1988, 1994 and 2000); average F_{ST} over this period for the five loci was 0.018 (p = 0.036)). That heterozygosity increased contrary to neutral expectations of drifting allele frequencies in a closed population, indicates that the period of time explored in the study encapsulated a natural fluctuation within a longer-term decrease in diversity. This study showed that despite relatively low genetic diversity relative to census numbers, roach populations were subject to the evolutionary force of genetic drift that significantly affected allele frequencies in a period less than 25 years.

4.1.6 Chapter Aims

The primary aim of this chapter of the thesis is to ascertain the levels of nuclear genetic (microsatellite DNA) variation in a common, generalist species of UK freshwater fish (the roach), applying classic and recent methodological analyses to answer questions of some import in fisheries management. Firstly, do roach from physically distinct river systems display significant levels of genetic divergence; and if they do, what can this tell us about the demographic or evolutionary forces that have taken place in approximately only 15000 years of independent evolution (time since post-glacial recolonization of northern European rivers) and 7500 years of demographic isolation (time since isolation of UK river systems from continental Europe by the English Channel/North Sea). Further, is a highly common species like roach worthy of being managed in a system that is generally the preserve of rare or commercially valuable species? Additionally, is there evidence for genetic divergence within the two study river systems (Thames and the Suffolk Stour); and if so, what is the nature of the divergence and can levels of genetic diversity be allied to equilibrium or non-equilibrium processes?

The first hypotheses that need addressing relate to population structuring between the Thames and the Stour. The Thames and the Stour are quite distinct rivers systems. The Stour possesses a much more uniform riverine habitat than the highly modified main Thames river. Due to larger population sizes in the larger Thames catchment, it may be expected that the Thames would maintain higher levels of genetic diversity than the Stour,

despite sharing a common historical breeding pool. The Stour's smaller size should result in fewer individuals, and thus an increase in inbreeding relative to that of the Thames. This comparison allows a test of the statistical criterion of management unit status, in that pure one-shot allelic frequency data is used. If positive, this would represent a significant result, in that in just over 7500 years (approximately 1875 generations) populations of a highly diverse, widespread and eurytopic fish species would have become genetically distinct enough to merit independent management status, despite a long history of translocation in the UK and western Europe.

Subsequently, structuring within rivers will be investigated to test a number of hypotheses relating to the rate and extent of migration among populations, averaged out over both juveniles (downstream drift) and adults (individually deterministic migrations). That juvenile drift occurs in roach as well as other coarse fishes is known (Peterka et al. 2004), with larger roach drifting greater distances than younger ones. There is, therefore, to be an expected correlation of genetic distance with geographic distance. However, continued return migrations to natal spawning sites should clearly partition some of the genetic variability into spatially delineated sub-populations, which may exacerbate or erode isolation by distance (IBD) dependent upon the number of spawning sites relative to the distances roach juveniles and larvae disperse. Previous studies have failed to find evidence for isolation by distance (Hänfling et al. 2004) or assumed panmicticism post hoc (Wolter 1999; Wolter et al. 2003). Therefore, a lack of IBD shall be the null hypothesis. Structuring may occur through non-equilibrium processes such as non-random mating and natural selection. Therefore, this section shall also investigate the likely correlations to be examined from available environmental data. Perhaps the most interesting of the latter concerns the levels of feminizing hormones in British waterways, which are known to cause some male roach to develop primary and secondary female sex characteristics. These intersex males have reduced fertility. The genetic consequences of feminizing hormones (EDCs, or endocrine disrupting chemicals) shall be of particular interest. Where the availability of breeding males is lowered, a greater skew in the local breeding effort is expected towards fewer successful males (reduced effective breeding size) and a concomitant reduction in local genetic diversity.

To investigate the nuclear microsatellite DNA variation of roach fully, the dataset must be assessed as being appropriate and fit for purpose. In section 4.2, the basic microsatellite genotype data are tested rigorously to see whether each microsatellite locus contributes an

independent point of information that is as free from bias as can reasonably be expected and within acceptable margins of error.

4.2 Sampling Scheme, Laboratory Methods and Data Quality & Assurance

4.2.1 Methods

4.2.1.1 Sampling

As with all roach collected for an appraisal of mtDNA variation, the professionals within the Fisheries Department of the UK's Environment Agency performed all physical sampling of roach. Their expertise is such that they can discriminate between all known roach *x* other cyprinid hybrids. The physical sampling procedure was the same as described in Chapter Three. All roach were sampled after the spawning period and after they had returned to summer feeding grounds. Thames roach were sampled from a time period of July – October 2005, whereas the Stour roach were sampled in a similar period a year later in 2006.

The primary aim of the microsatellite study is to investigate the levels of genetic structuring within linear stretches of river. See Figure 4.1 for a map of each of the rivers and of the sampling sites along their lengths, and Table 4.2 for all sampling details, including specific geographic coordinates and the numbers of fish sampled per location. All site names are taken from the nearest mill, lock or weir directly upstream from where roach were sampled, and account for roach fished from within an approximate 100m stretch of river downstream, or until the next sampled site at a downstream mill, lock or weir. Therefore, for both the Stour and the Thames, sampling sites were optimized to be as far apart as possible between the upstream reaches of roach habitation and the downstream tidal halocline. Sometimes, however, the lack of a suitable number of individual roach present at optimal distances meant either sampling a short distance upstream or downstream, or pooling samples where otherwise there would be a significant gap between sites (e.g. in the Thames data the population at 'Eynsham' pooled the Eynsham roach with those roach individuals found downstream at Odney Weir, whereas in the Stour

individuals were pooled to make up the Stoke-by-Clare population). However, pooling data was kept to a minimum, and the minimal number of roach per sample was maintained at 20 individuals per population, although the mean numbers of individuals per site is much greater (see Table 4.2).

Table 4.2: Sampling details for each location at which fish were caught for genetic analysis

Sampling Site	Code	Geographical Coordinates		Sample Number
		Latitude	Longitude	
Thames				
Molesey Weir Pool	MWP	51.405637	-0.345167	33
Desborough Loop	DL	51.383549	-0.439374	65
Old Windsor	OW	51.485767	-0.589391	49
Clivedon Island	CI	51.545816	-0.693418	63
Temple	Te	51.551675	-0.792194	60
Whitchurch	Wh	51.486617	-1.089740	23
Dorchester	Do	51.641830	-1.164674	33
Days	Days	51.638345	-1.180634	33
Culham	Cu	51.646155	-1.274436	40
Eynsham*	Ey	51.775208	-1.356433	33
Northmoor	No	51.716871	-1.376079	26
Buscot	Bu	51.681196	-1.668736	27
Roundhouse	Ro	51.686687	-1.704859	22
				Σ 507
Suffolk Stour				
Brantham Lock	BL	51.956858	1.040325	86
Denham Mill	DM	51.963634	0.995650	40
Stratford Weir**	St	51.961420	0.976577	84
Anchor Bridge	AB	51.968510	0.872212	59
Shalford Weir	Sh	52.007552	0.743557	64
Mill Meadow	MM	52.038694	0.719216	78
Rat's Castle	RC	52.078182	0.603393	20
Stoke-by-Clare***	SbC	52.057994	0.539488	43
Thurlow	Th	52.122522	0.455658	20
				Σ 494

*Eynsham includes individuals sampled from the adjacent weir downstream, Odney; ** Stratford Weir includes samples just upstream of Denham Mill; *** Stoke-by-Clare includes roach from the adjacent downstream site of Mill Green.

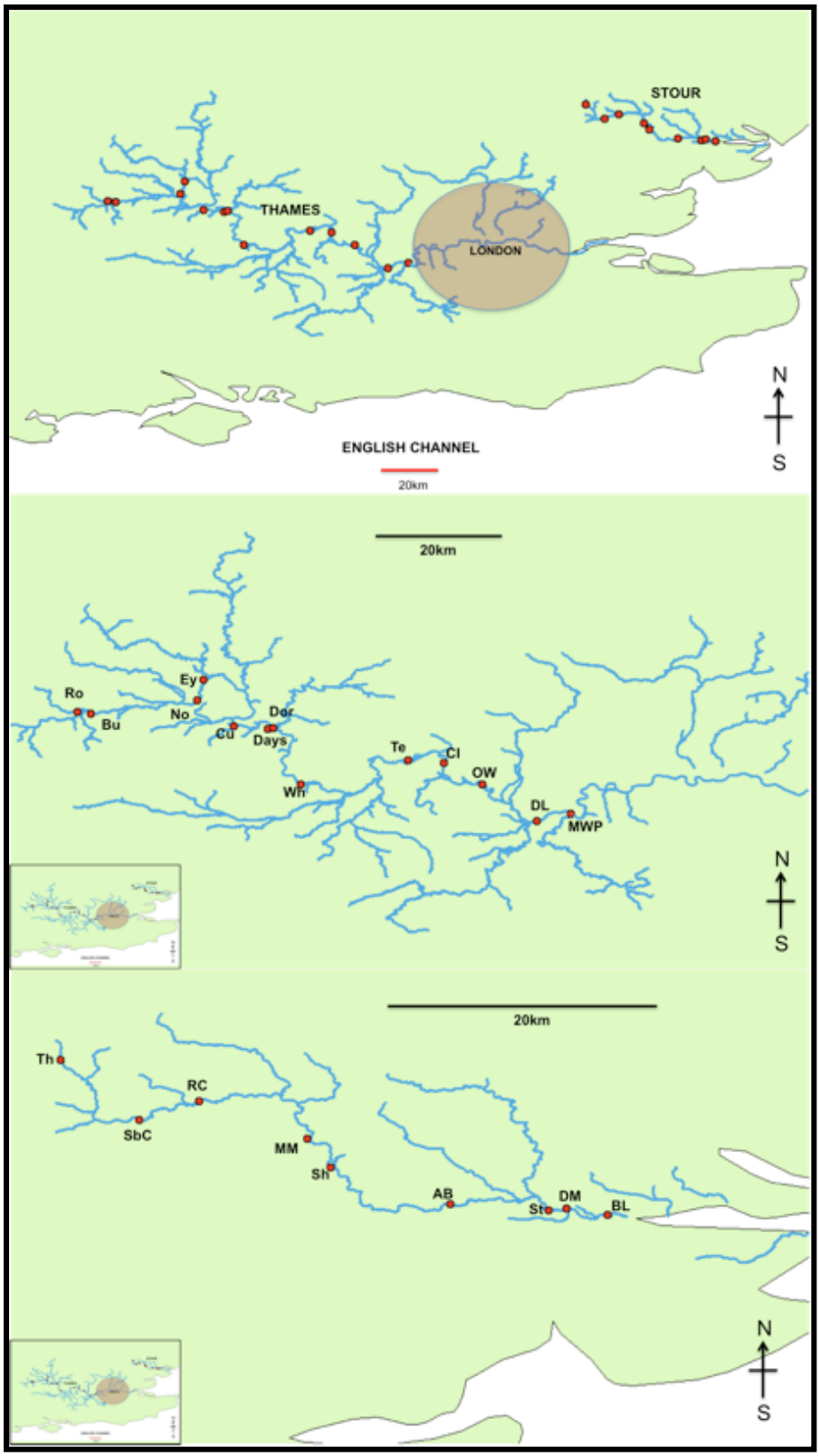


Figure 4.1: Map of sites at which roach were sampled for microsatellite variation in the SE of England. Top panel: Thames and Stour watersheds in geographical context; Middle panel: Thames; Bottom panel: Suffolk Stour. See Table 4.2 for location code details.

Because the stretches of the rivers being compared are variable in length (from source to mouth: Thames is 185.04km; Stour is 65.6km), the Thames dataset necessarily needed more samples located along its length to test for IDB (13 populations as opposed to 9 in the Stour). Similarly, too small a sampling scale and one might sample the same population multiple times. However, without *a priori* knowledge of putative populations one cannot rule out the possibility of sampling the same deme more than once. This is more likely to be a problem in the Stour than in the larger Thames. Additionally, the location-specific migratory behavior of roach - reproductive, diel and winter - in the two rivers is unknown. The sampling scheme for IBD is largely a compromise between practical realities (numbers of roach caught and the location where the school of fish was at the moment of capture) and statistical requirement.

4.2.1.2 Laboratory methods

4.2.1.2.1 DNA extraction

Total DNA (i.e. including nuclear DNA) was extracted using the same protocol set out for mitochondrial DNA (see Chapter Three for details).

4.2.1.2.2 PCR Conditions: Locus Derivation and PCR Optimization.

Each of the microsatellite loci used in this project were derived from published studies of the efficacy and efficiency of cross-amplification of cyprinid loci in other confamilial species. Table 4.3 lists the names, the repeat motif, the number of published alleles, the PCR-product size range and the publication of origin for each locus. Each locus was selected for the potential to yield a high number of alleles per locus and for a high probability of polymorphism across populations. Trial runs for a number of published candidate loci were carried out across a selection of populations to determine the locus set to be screened.

Table 4.3 Microsatellite locus data for population genetic analyses of the roach

Locus	Repeat motif	Original publication	No. alleles	Size range of alleles
Rru3	$(ACTC)_5N_{21}(GT)_7A(TG)_6$	[1]Barinova et al. 2004	4[1]; 5[2]	176-190[1]; 169-179[2]
Lid1	$(CT)_5(CA)_{20}$	[1]Barinova et al. 2004	4[1]; 8[2]	236-274[1]; 248-285[2]
CypG3	$(CAGA)_2(TAGA)_{11}$	Baerwald & May 2004	16 [2]	194-342[2]
CypG48	$(TAGA)_8TACGG(TAGA)_{10}$	Baerwald & May 2004	-	-
Ca1	$(CA)_{24}$	Dimsoski et al. 2000	7 [2]	103-137[2]; 106-118[4]
Ca3	$(TAGA)_{14}$	Dimsoski et al. 2000	18 [2]	240-308[2]; 275-303[4]
Ca12	$(TAGA)_{10}(CAGA)_4(TAGA)_2$	Dimsoski et al. 2000	-	175-243 [4]
Lc27	$(CT)_{22}(CACT)_3(CT)_2$	[3]Vyskočilova et al. 2007	3 [3]	144-152[3]
Lc290	$(GA)_4N_{49}(CT)_{13}TT(CT)_{15}CC(CT)_2CC(CT)_{11}CC(CT)_3$	[3]Vyskočilova et al. 2007	6 [3]	178-198[3]
Lco4	$(GT)_5ATTTT(GT)_5(GA)_{11}$	Turner et al. 2004	2 [3]	226-228[3]

[2] Hamilton & Tyler 2008; [4] Holmen et al. 2005.

All amplification reactions were carried out according to published conditions for each locus, although all reactions were standardized to the following volume: a total 10μl reaction volume composed of approximately 50ng of genomic DNA, 0.2 units of Taq DNA polymerase (Bioline 5U μl^{-1}), 1.5mM $MgCl_2$, 0.125mM of each of the four deoxynucleotide triphosphate (dNTPs, also Bioline) nucleic acid monomers, 0.2 μM of the 3'- unlabeled and the 5'- end-labeled (with fluorescent dye) primers and 1x PCR Buffer. The remainder of the volume was made up of double-distilled H_20.

The visualization of PCR products was carried out on 6% polyacrylamide gels within ALFexpress II and III^{TM} (Amersham Pharmacia Biotech) automated sequencers, alongside a molecular ladder of known size fragments. By using proprietary software to size-quantify each allelic PCR-fragment (Fragment Manager version 1.2) against the known fragment ladder (consisting of three or four fragments of known length, which may be altered for differently sized microsatellite loci), each individual roach was available for genotyping at each of the ten loci.

4.2.2 Data Quality & Applicability

4.2.2.1 Statistical Independence of Genetic Markers.

4.2.2.1.1 Rationale

The most basic requirement for population genetic studies reliant upon information attained by the use of Mendelian co-dominant markers is for all genetic markers to be inherited independently, according to Mendel's second law of independent assortment. If this were not the case, then each locus would not contribute an independent point-estimate of genetic diversity and differentiation. Therefore, any estimates of any genetic parameters computed from such a dataset would be statistically compromised. The non-independent statistical association of allelic classes across loci in all sampled populations, at any sampled temporal period, is strong evidence that these nominal loci are *physically linked*. If any loci are found to be linked then all but one of the inferred physically linked loci – normally the most polymorphic – are discarded from the study. Linkage disequilibrium analysis was conducted in FSTAT version 2.9.3.2 (Goudet 1995), analyzing all 1001 pooled individuals. If disequilibrium is not found to be significant when all individuals are pooled then all loci may be considered independent Mendelian markers. However, if there is significant association of allelic classes among some loci, then further analysis is needed to discount the effect of population admixture[29] in determining non-independent associations. All 45 possible locus-by-locus associations were tested for significance at the alpha level of 0.05 after permutation (corrected for multiple comparisons using the Bonferroni method (Rice 1989)).

4.2.2.1.2 Results

After a round of 900 permutations, the results of the linkage disequilibrium analysis of the pooled locus data suggest that none of the loci are physically linked. However, one locus-by-locus association was borderline significant before the significance level was adjusted for multiple comparisons (p = 0.0011, Bonferroni α = 0.0011). However, when the individual rivers were tested separately, this association was no longer present.

[29] Inbreeding, associative mating and natural selection may also bring about statistically significant associations among alleles at unlinked loci, but the Wahlund effect is most probable at such a scale.

4.2.2.1.3 Conclusion

One can conclude with some confidence that each of the ten loci are located on separate chromosomes, or are located on the same chromosome but at such a distance that recombination between them is complete, and that each locus contributes an independent, Mendelian source of information to subsequent analyses.

4.2.2.2 Null Alleles: Presence and Influence

4.2.2.2.1 Rationale

In addition to determining the independence of the study loci, one must uncover whether the genetic variation at a particular locus is influenced to a significant degree by cryptic variation and/or the laboratory-based failure of allelic amplification during PCR. In other words, one must guard against skewed frequencies of any one allele, at any one locus, due to the presence of null alleles, whose presence within a dataset is due either to 'natural' misidentification of heterozygotes as homozygotes because of the occurrence of mutations in primer-annealing sites (Angers and Bernatchez 1997; Callen et al. 1993) and/or large allele drop-out (Wattier et al. 1998); or through human error associated with the allele-scoring process (laboratory error, poor DNA quality or allelic-stuttering of PCR-product in acrylamide gels).

The presence of null alleles may directly impact upon estimates of population differentiation and other such analyses dependent upon the calculation of allele frequencies (Dakin and Avise 2004). Null alleles will tend to over-inflate estimates of genetic divergence between loci in different populations (Pemberton et al. 1995). Directional selection at, or near, a locus may also deviate that genomic region from HWE (Dakin and Avise 2004). Further, whether null alleles appreciatively affect genetic parameter estimates is not fully resolved, with no clear consensus in the literature. Chapuis and Estoup (2007) define a high frequency of null alleles as being ≥ 0.2, and moderate frequencies accounting for between 0.05 and 0.02 of all allele frequencies for a particular locus and population. A general rule of thumb, as a minimal requirement for acceptance in a population genetic study, should be to estimate parameters (such as F_{ST} and other measures of genetic distance) using both unadjusted and adjusted allele frequencies, as can be determined by such programs as FreeNA (Chapuis and Estoup 2007). If the difference

across all loci is statistically significant, then appropriate measures should be taken (e.g. dropping affected loci from the study).

Null allele frequencies were estimated using the FreeNA software of Chapuis and Estoup (2007). FreeNA is a program that utilizes an Expectation Maximization (EM) algorithm (Dempster et al. 1977) to estimate the frequencies of null alleles per locus. Other estimators are available, however the EM of Dempster et al. performs better, at least in a comparative setting. FreeNA is less susceptible to bias and it calculates smaller variances (Chapuis and Estoup 2007). This analysis was conducted on the dataset as a whole (all 22 sampled 'populations' pooled as a single set of 22 populations) and on a partitioned dataset of both the Stour and the Thames populations (n = 9 and 13 populations, respectively). For each analysis the data was permuted 1000 times to determine statistical significance and 95% confidence intervals. Unbiased F_{ST} (Weir's estimator (1996)) was estimated for the degree of allelic divergence between the Stour and the Thames both before and after correcting for the influence of null alleles using the ENA method.

4.2.2.2.2 Results

Table 4.4 reveals the results of the EM null allele frequency analysis. Null allele frequencies were classified as negligible, moderate or large according to the definition of Chapuis and Estoup: frequencies below 0.05 are deemed negligible; those between 0.05 and below 0.2 are labeled as moderate, whereas frequencies greater or equal to 0.2 are considered to be large. Shown in the far left column of Table 4.4 are the three categories of null allele frequency and their allocated frequency parameter, *r*. For the pooled (global) dataset and for the Stour and Thames populations, the frequencies of putative null alleles for each of the possible locus-by-population comparisons are apportioned in each of the three categories. These figures are also shown as a percentage of the total population dataset.

The vast majority of putative null alleles fall within a frequency bracket no greater than moderate in strength for the global population data, and for each partitioned river dataset. Analyzing over all 22 populations suggests that no single hypothesized null allele would reach a level as to significantly impact upon estimates of genetic parameters that are dependent upon robust allele frequency data (35% of putative null alleles are negligible in frequency, whereas 65% are of moderate significance).

Table 4.4. Expectation Maximization analysis of null allele frequency and severity

Category	*r*	Global		Stour		Thames	
		Frq.	***%***	***Frq.***	***%***	***Frq.***	***%***
Negligible	$r \leq 0.05$	7	35	49	54.44	78	60.00
Moderate	$0.05 \leq r < 0.02$	13	65	37	41.11	51	39.23
Large	$r \geq 0.02$	0	0	3	3.33	0	0
Missing Data	-	0	0	1	1.12	1	0.77
	Σ	20	100	90	100	130	100

Only within the Stour populations, when the Stour is analyzed separately, are significant null allele interactions observed (3 locus-by-location comparisons out of 90 possible combinations, or 3.33% of all locus-by-location combinations). Large null allele effects are limited to just two loci within the Stour: Lco4 and Rru3, and only in the following populations: Stoke by Clare (locus: Lco4), Denham Mill (locus: Rru3) and Anchor Bridge (locus: Rru3). Furthermore, the proportion of negligible null allele frequencies over each combination is greater than the proportion of moderate null alleles in the partitioned datasets than is the situation in the global dataset (negligible r = 35%, 49% and 60% for the global, Stour and Thames data, respectively; moderate r = 65%, 41.11% and 39.23% for the global, Stour and Thames data, respectively).

Although there is little evidence for widespread instances of severe null allele frequencies, moderate levels may still exert a significant bias upon subsequent investigations of population structuring and genetic diversity. The level of genetic differentiation between the Stour and Thames was measured using Weir's (1986) estimator of the F_{ST} statistic and was calculated both prior to, and after, the incorporation of putative null allele frequency information, by implementing the ENA procedure in FreeNA. The results of this analysis for each of the ten loci and over all loci are shown in Table 4.5. F_{ST}, assuming that all allelic diversity is present in the sample data, between the Stour and Thames is given as 0.0205 (95% CI: 0.0106 – 0.0317) (the discussion of structuring between these two rivers is in section 4.3), whereas after accounting for some error associated with the over representation of homozygotes, F_{ST} was determined to be 0.0192 (95% CI: 0.0097 – 0.0301). As expected, if one assumes some overrepresentation of homozygosity caused by mis-assigned heterozygotes, the ENA-adjusted F_{ST} scores for all loci are lower than the uncorrected F_{ST}s. The relationship between F_{ST} and ENA-F_{ST} should be perfectly linear if all loci are unaffected significantly by null alleles. Figure 4.2 shows just such a linear

regression between F_{ST} and ENA-F_{ST} (performed in Mathematica version 8.0). The line of best fit (blue line in Fig 4.2) is less than one, indicating some deviation from perfect predictability, but the predictive relationship between F_{ST} and ENA-F_{ST} is not significantly affected (2-tailed $p << 0.001$), thus putative null alleles minimally affect estimates of F_{ST}.

Table 4.5 Estimates of F_{ST} (Weir (1996)) between the Thames and the Stour before and after correction for putative allele frequencies at all study loci and over all loci

Locus	F_{ST}	F_{ST}-ENA
Lid1	0.0073	0.0067
Lco4	0.0513	0.0486
Rru3	0.0371	0.0342
Ca1	0.0095	0.0086
Ca3	0.0028	0.0025
Ca12	0.0125	0.0124
CypG3	0.0271	0.0266
CypG48	0.0068	0.0065
Lc27	0.0041	0.0020
Lc290	0.0388	0.0373
All	0.0205	0.0192
2.5%CI	0.0106	0.0097
97.5%CI	0.0317	0.0301

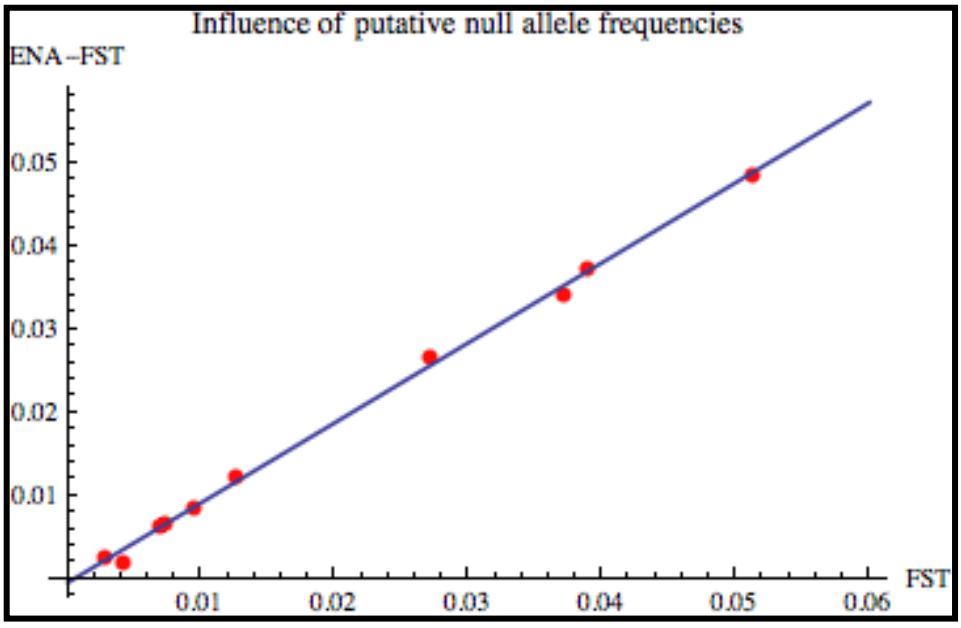


Figure 4.2 Linear Regression of F_{ST} (x-axis) with ENA-F_{ST} (y-axis).

4.2.2.2.3 Discussion

Null alleles are problematic for population genetic analysis and may be present in increased numbers in taxa with high effective population sizes. This variability is hypothesized to be caused by the inherent instability of microsatellite flanking sequences as compared to other areas of the genome, and not due to the number of repeat units or the complexity of the repeat motif itself (Chapuis and Estoup 2007). If null alleles are common in such species, then the assumptions of most corrective analyses (that a single null allele per locus[30] is present in all sampled populations (Roques et al. 1999)) are probably inaccurate (Paetkau and Strobeck 1995; Chapuis and Estoup 2007). Nine of the ten microsatellite loci employed in this study were derived from cyprinid fish outside of the *Rutilus* genus, but from within the leuciscinid subfamily. Due to the evolutionary distances among taxa within this subfamily, it should be unsurprising that mutations may have occurred within the primer-binding sites in the millions of years since divergence from a common ancestor. Instances of the non-amplification of microsatellites using primers developed in an original model species have been recorded (Dakin and Avise 2004). The locus that is most suggestive of null interference in the current dataset is the Rru3 locus (in which the allele '179' was overrepresented in numerous homozygotes, especially in two Stour populations (Denham Mill and Stoke by Clare)), which was originally sourced from a specimen of *Rutilus rutilus*. However, the specimen was a Caspian roach, a putative separate sister taxon to *R. Rutilus* (Ketmaier et al. 2008, and see Chapter Three).

Whether null allele frequencies have the potential to seriously bias estimates of population differentiation depends upon the underlying demographic properties of the system under study. Simulation studies suggest that in scenarios in which gene flow among populations is high then the bias caused by null alleles is expected to be low (Chapuis and Estoup 2007). Such an *a priori* expectation may be apt for roach populations that are subject to unidimensional channels of migration in the face of high fluvial rates of flow. If roach do have a preponderance of null alleles, which is not significantly inferred by the EM-analysis of the data, then contemporary high gene flow may go some way to minimizing their effect upon measures of population differentiation, at least for estimates of divergence within rivers. Assuming a base-pair substitution rate of 10^{-9} per nucleotide site, Chapuis and Estoup calculate the probability of mutation occurring in the latter 10bp sequence in the 3' flanking region of a 20bp oligonucleotide primer pair to be 2×10^{-8}.

[30] That is to say that the null allele is associated with only a single repeat-motif size class.

This translates into an estimate that null alleles are only likely to be found in populations whose effective population sizes exceed 50000 individuals. As the authors concede, null alleles have been observed in species with effective sizes much less than this theoretical estimate. However, if many microsatellite primers are adopted in species in which they were not developed, then the evolutionary distances between the original and adopted species may constitute very large evolutionary effective population sizes, thus accounting for some of the discrepancy between theoretical and inferred levels of null alleles in surveyed taxa.

The above EM-analysis assumes all loci to be in HWE. However, the Wahlund effect produces similar genetic signals (e.g. heterozygote deficits) to those resulting from the presence of null alleles. The effect of null alleles and deviations from HWE is likely to be a synergistic relationship, as they both result in the overrepresentation of homozygotes (although population structuring and the sampling of admixed populations should result in Hardy-Weinberg disequilibrium across all or most loci for particular locations). The results of the EM analysis then present the 'worst-case scenario', in which both null alleles and admixed and/or inbred (Pemberton et al. 1995) and/or naturally or sexually selected populations (Dakin & Avise 2004) contribute[31]. Thus, even if there is significant Hardy-Weinberg disequilibrium and other contributing factors, the influence of null alleles is still determined to be weak as the potentially synergistic influences of all sources of heterozygote deficit are implicit in the data.

4.2.2.2.4 Conclusions

Given the seemingly weak influence of null alleles (whose signal may be conflated with other population genetic causes), in addition to the non-significant difference in F_{ST} estimation, coupled with known levels of roach genetic diversity and inferred migration from other genetic studies, the occurrence of null alleles was not considered a significant source of error for all ensuing genetic analyses.

[31] One has no *a priori* information regarding whether the sampled populations are true, outbreeding demes, and for now the question of whether the loci are in HWE or potentially influenced by genetic hitchhiking is ignored.

4.2.3 Deviation from Neutrality

4.2.3.1 Rationale

Genetic structuring among populations based upon the surveying of neutral marker loci depends entirely on their selective neutrality, if one is to test hypotheses based upon deviations from mutation-drift-migration equilibrium. If one is to base estimates of population genetic divergence upon an underlying assumption about the lack of migration, then one has to rule out from subsequent analyses any loci that show orthogonal levels of genetic variation when compared to the loci-set as a whole. That is to say, outlier loci, when measured for their contribution to genic differentiation, may be physically linked to neighbouring genomic areas under the influence of natural selection. Such loci, if their frequencies were taken at face value (e.g. large variances among allelic classes, i.e. larger than average F_{ST}), would suggest that migration estimates also based on these frequencies would be low, when in reality they would be higher than the data would suggest. Such loci would violate the assumptions of marker neutrality. Whilst estimates of divergence would still be informative (e.g. one may check for environmental correlations with particular allele frequencies) in terms of structuring, it would be necessary to discard such loci from analyses that assume selective neutrality (e.g. all estimates of migration).

There are a number of programs available that scan genomic regions for deviations from strict neutrality. Each of these methods endorse the theory that loci under directional selection will tend to show greater levels of genetic divergence among populations and reduced amounts of genetic diversity within populations than loci behaving neutrally (Lewontin & Krakauer 1973). Furthermore, under positive selection the effect is local, whereas neutral processes affect all neutral markers around the genome with equal probability. Firstly, a coalescent model-based procedure that accounts for population subdivision was implemented in the LOSITAN workbench (Antao et al. 2008), incorporating the Fdist program of Beaumont and Nichols (1996**).** This procedure permutes the expected relationship between H_e (expected heterozygosity) and F_{ST}, assuming Wright's n-island model of population structure (equal population sizes exchanging symmetric amounts of migrants), and migration-drift equilibrium. LOSITAN resamples the distribution of H_e and F_{ST} (the data were resampled 10^5 times). The dataset was run twice assuming both the IAM and SSM models of microsatellite evolution. The second method, implemented in the software BAYESFST (Balding 2003**;** Beaumont &

Balding 2004**),** uses hierarchical Bayesian inference to determine the probability that each marker locus possesses F_{ST} values that lie outside the 95% interval of a normal distribution. Non-neutrality is inferred if the locus resides in either the lower or upper quartile after summing the results of the simulations. This latter method also takes into account different sample sizes, unlike Fdist. BAYESFST was run separately on the Stour and Thames datasets to test whether sample size plays some part in contributing to the signal of outlying loci, whereas all 22 populations and the Stour and Thames populations were tested by Fdist. In BAYESFST, the MCMC algorithm was employed to generate the posterior probability distribution from which 2000 draws were made and analysed in the statistical package R 2.9.1 (http\\CRAN.R-project.org).

4.2.3.2. Results

The results of the Fdist analysis are displayed graphically in Figure 4.3. Shown are plots of actual F_{ST} versus expected heterozygosity for each locus for datasets comprising all 22 populations, in addition to the 9 and 13 populations of the Stour and Thames, respectively. The coloured areas represent the theoretical zones within which certain combinations of simulated F_{ST} and gene diversities indicate whether loci resident in these zones are neutral with respect to selection (grey area), under the influence of directional selection (red area) or under the influence of balancing selection (light blue area). Because the theoretical expectations of the SMM and IAM models did not differ greatly, only the SMM results are shown[32]. When the dataset is partitioned into sampled populations, the simulated neutral zone is heavily constricted in the entire collection of 22 populations and in the Thames, such that most loci are presumed to be under some sort of selective regime.

[32] Strict SMM and IAM probably do not best reflect the reality of microsatellite evolution in the roach, with the strict SMM model rejected as a possible influence upon estimates of genetic divergence (see section 4.3.4.2). Two-phase models (Di Rienzo et al. 1994) provide a better approximation in these scenarios, but such a model is unavailable in LOSITAN.

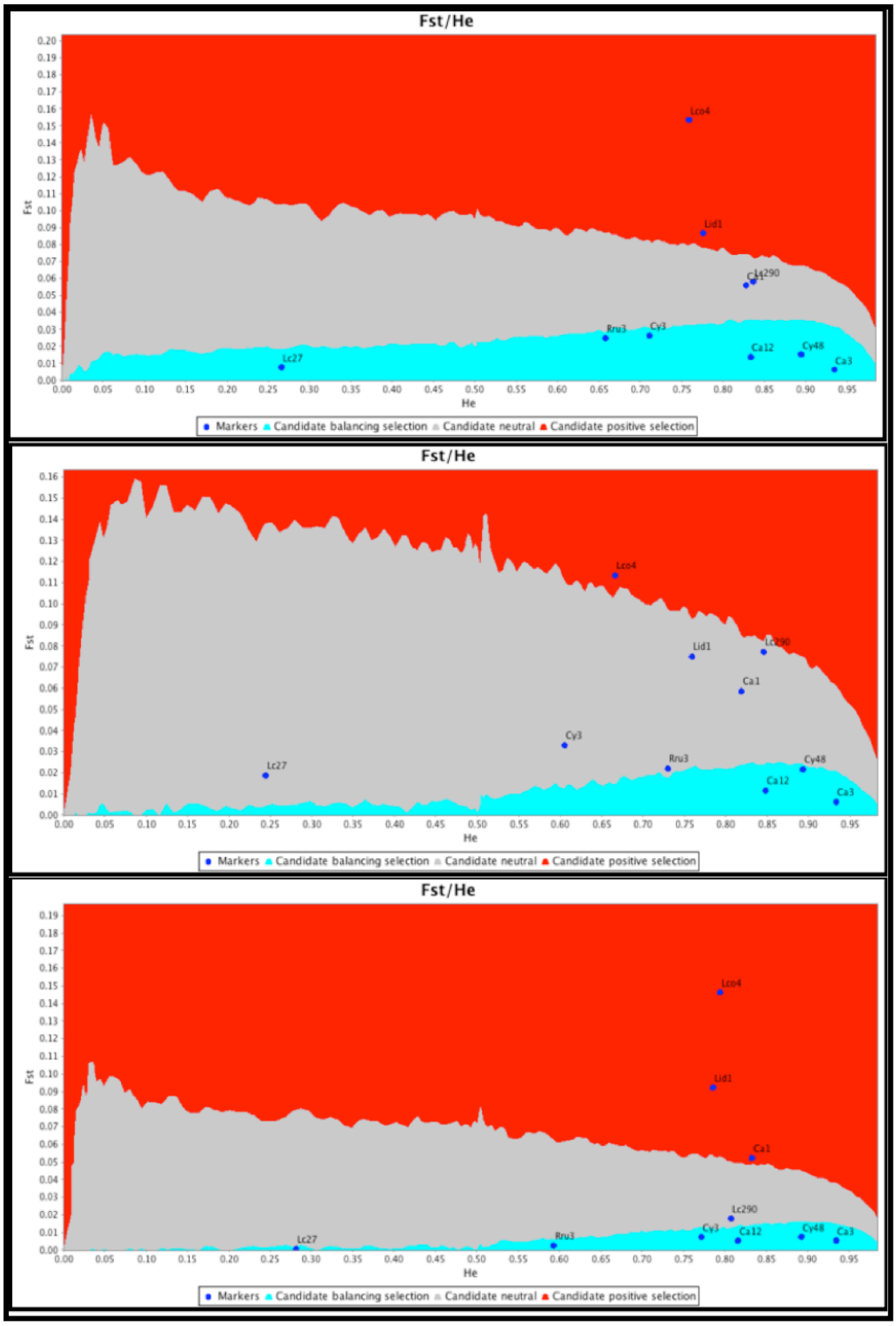


Figure 4.3: Results of the Fdist analysis as displayed within the desktop LOSITAN workbench: Upper panel: 22 populations; Middle panel: Stour populations; Lower panel: Thames populations. All analyses were conducted assuming the SMM model of microsatellite evolution.

Only one locus, Lco4, is found in the red zone of directional selection in all three population-level analyses (probability that simulated $F_{ST} < F_{ST} = 1.0$, 0.983, and 1.0 in the 22-population, Stour and Thames populations, respectively). From studying the graphs, it is clear that the greatest level of deviation from neutrality is found in the Thames, and this contributes to the deviatory nature of the global analysis.

Further, BAYESFST was employed as an ancillary analysis. Fig 4.4 shows the relationship between genetic divergence (F_{ST}) with the log-transformed *p*-values of the hierarchical Bayesian relationship between locus-specific, population-specific and interaction effects, for each locus and summarized over all populations for a) the Stour and b) the Thames. Loci (dots) shown in red are those that behave in a non-neutral manner according to the Fdist analysis above, whilst the vertical line represents the critical cut-off value for statistical significance ($\alpha = 0.05$). The vertical line, superimposed on the graph, represents the statistical confidence limit, such that red loci to the right of the line are statistically significant ($p < 0.05$). Locus Lco4 exhibits a strong, positive relationship between locus-specific and population-specific effects, suggestive of the influence of positive selection. It is the only locus, besides Lid1 (in the Thames), to remain significant after the critical cut-off. No other locus was significantly associated with deviations from selective neutrality.

4.2.3.3 Conclusions

Because Lco4 appears to consistently violate expectations of selective neutrality, even after controlling for the influence of sample size, it was removed from all analyses in which further genetic or demographic processes are inferred (except for environmental correlations). Also, erring on the side of caution, Lid1 was also used only for structuring and correlative analyses, and not for inferring migration.

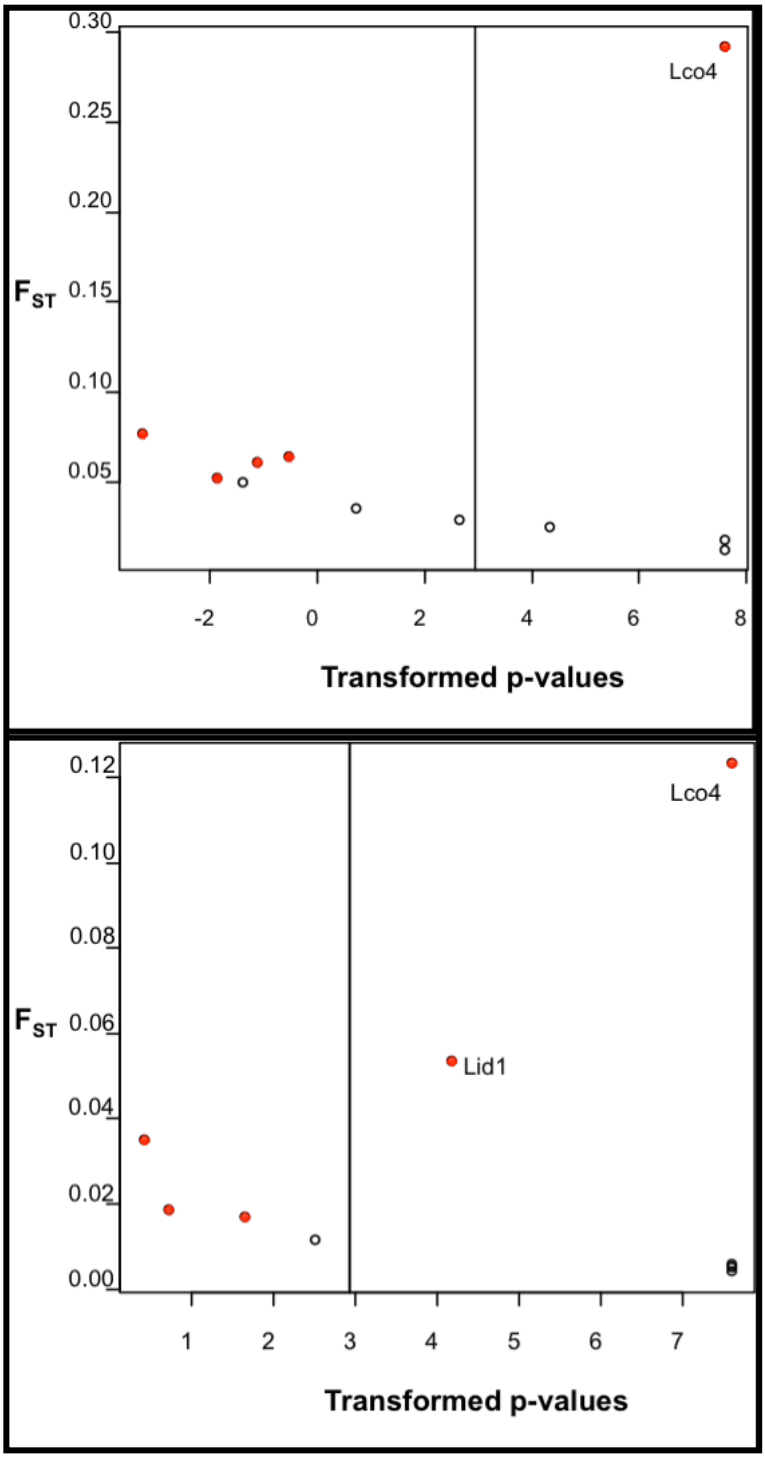


Figure 4.4: Graphical representation of the BAYESFST analysis for both the Stour (upper panel) and the Thames (lower panel) populations.

4.2.4 Levels of Microsatellite Variation

4.2.4.1 Rationale

The degree to which microsatellite loci exhibit high levels of allelic variation is largely dependent upon the effective population size of the organism in question, coupled with the mutation rate per generation of the loci themselves. Previous studies of the roach have found a close correlation between genetic diversity (as measured by gene diversity) and high census numbers for a range of nuclear markers, including allozymes (Bouvet et al. 1991, 1995; Baranyi et al. 1997; and Wolter 1999), RFLPs (Wolter et al. 2003) and microsatellites (Demandt & Björkland 2007; Demandt 2010). Therefore, it was expected that UK roach would possess similarly high levels of variation (if not higher given the sample numbers and nature of the loci). High levels of genetic variation may increase statistical power to detect significant differentiation of populations via differential allele frequencies in those populations. Numerous loci were tested to gauge levels of genetic diversity and polymorphism to screen them for utility in population genetic analysis. Diversity estimates, including the total number of alleles, the number of effective alleles, allelic richness, observed and expected levels of heterozygosity were conducted using FSTAT for a number of candidate microsatellite loci, to determine those best suited for the population genetic analysis of the common roach. The diversity indices of the ten loci used in this study are expanded below.

4.2.4.2 Results

Table 4.6 shows the results of the diversity listed above, for all loci individually, and as means and standard errors for each locus. The mean number of alleles over all ten loci is 20.5, which translates to an effective number of alleles of 5.730. Accordingly, due to the relatively high number of alleles, both allelic richness and expected heterozygosity are also consonantly high (mean A_r = 20.134 based upon a minimum of 767 individuals, H_e = 0.740). The data contain a low number of observed heterozygotes (all loci violate HWE, $p << 0.001$), which was anticipated due to the pooling of two presumed independent, interbreeding populations.

Table 4.6: Genetic diversity estimates for ten microsatellite loci

Locus	NGI	NA	NEA	A_r	H_o	H_e
Lid1	941	15.000	4.190	14.596	0.599	0.761
Lco4	967	10.000	3.466	9.749	0.555	0.711
Rru3	873	14.000	3.033	13.698	0.464	0.670
Ca1	872	20.000	5.582	19.745	0.662	0.821
Ca3	808	32.000	15.079	31.746	0.816	0.934
Ca12	958	30.000	6.378	29.163	0.756	0.843
CypG3	872	37.000	3.400	36.097	0.622	0.706
CypG48	767	22.000	9.323	22.000	0.816	0.893
Lc27	975	8.000	1.326	7.739	0.223	0.246
Lc290	941	17.000	5.521	16.807	0.672	0.819
Mean	897.40	20.500	5.730	20.134	0.618	0.740
SE	22.416	3.070	1.249	3.018	0.056	0.061

NGI = number of genotyped individuals; NA = number of alleles; NEA = number of effective alleles; Ar = allelic richness; H_o = observed heterozygosity; H_e = expected heterozygosity.

The number of alleles ranges from 8 (Lc27) through to 37 (Cy3), a spread that should enable significant testing of population structuring. The effective number of alleles portrays the data in terms of the proportion of allele frequencies in a given population. Lc27 has an NEA of 1.326, because one or two alleles dominate in the pooled population, whereas Cy3 has an NEA of 15.079 because there is a greater spread of alleles at or around equal frequency. Some investigators use NEA to calculate levels of population differentiation.[33]

4.2.4.3 Conclusions

The data appears to be free from the widespread influence of null alleles; loci are also inherited in a manner consistent with Mendel's second law, vital for statistical independence; and whilst some loci (Lco4, Lid1) may be best omitted from estimates of migration due to indication of the effects of selection, each locus is variable enough and sufficiently powerful to uncover population structuring if it exists

[33] Naglyaki (1985) showed that if populations are subdivided into finite demes, then in the absence of migration, demes diverge such that values of NEA should exceed their values expected in a panmictic population. Thus NEA can be used to determine population subdivision.

4.3 Testing for Genetic Differentiation of the Roach Populations of the Rivers Thames and Suffolk Stour: Implications for MU Designation in a Coarse Fishery

4.3.1 Introduction

Moritz (1994) succinctly defined genetically differentiated populations that are suitable for individual management consideration as those that exhibit a statistically significant deviation in their allele frequencies. Such designation of population structure is based upon a 'statistical criterion' (Pasbøll et al. 2006, 2010). However, statistical significance is implicitly defined by the quality and quantity of samples (e.g. Mace 1964) and thus statistical analyses, such as the determination of differences in allele frequencies, are in turn affected by this issue (e.g. both the numbers of loci and the number of individuals that are sampled affect significance levels) (Waples & Gaggiotti 2006; Pasbøll et al. 2010), including sampling from a limited distribution of real genomes and pooling individuals without any prior consideration of their derivation. One notable weakness of defining genetic populations based upon levels of allelic differentiation among them, is that ancestral levels of genetic variation may be so considerable in extent that any subsequent sub-division of the parent population into daughter populations may retain similar levels of genetic variation, but are by all intents and purposes genetically isolated and demographically independent, even after a significant number of generations have elapsed.

Large stable populations of breeding individuals are predicted to possess greater levels of genetic diversity than smaller populations more vulnerable to stochastic environmental and demographic processes. This assertion, whilst generally true, is also a function of various ecological and demographic considerations such as habitat availability, population fragmentation, local predator-prey assemblages and interactions, resource availability and inter- and intra-specific competition, in addition to localized stochastic and deterministic events. All these factors will affect growth, maturation and/or fecundity, which may positively or negatively influence the extent of spatially and temporally distributed genetic diversity. Estimates of genetic diversity also provide a baseline assessment of the degree to which genetic diversity is apportioned within and among populations within a species. From such measurements, axiomatic divergence estimates, which have been the building block of most population genetic studies over the past four decades, are derived.

Larger river systems are potentially greater repositories of genetic diversity within potamodromous species than are smaller, less dendritic systems, for a variety of reasons. Larger systems may flow through a wider variety of geologically variant topsoils and bedrock, which will influence riverine habitat types, as well as the mineral composition of the stream water itself. Genetic diversity has been positively correlated with higher levels of habitat heterogeneity and environmental (spatial and temporal) flux (Hedrick et al. 1976; Hedrick 1986) for a number of taxa, including freshwater fish species and communities (Vrijenhoek 1979; Smith et al. 1983; Coelho & Zalewski 1995; and Blum et al. 2011). Habitat heterogeneity may manifest greater levels of genetic diversity through differential selection in different environments (Hedrick 1986), or through frequency-dependent selection in the absence of over-dominance at any particular loci, thereby increasing overall levels of polymorphism (Levene 1953). Life-history variation may also affect the amount of genetic diversity within species, and between populations if the populations differ in this respect, as rates of demographic increase or decrease are dependent upon fecundity, time to maturation and death rates (Mitton & Lewis 1989).

Neutral models of differential genetic diversity between populations or species are also dependent upon demography. Larger habitable rivers would allow more individuals to survive, with a concomitant increase in the number of breeding individuals (i.e. higher N_e = reduced influence of genetic drift, which erodes heterozygosity at a rate less than would occur in smaller population sizes (Soulé 1976)). Temporally stable populations would also be expected to be more diverse than less stable populations, as younger populations, or those which are expanding after a recent bottleneck, are not maintaining diversity as efficiently as those at equilibrium; as genetic drift is the predominant evolutionary force post-expansion (Soulé 1972). Larger river systems may also provide local refugia for many individuals when their wider ranges are in some way compromised (Sedell et al. 1990).

4.3.2 Hypotheses

4.3.2.1 Levels of Genetic Diversity in Two Differently Sized Rivers

Assuming equilibrium conditions are met in the two rivers, one would expect a positive correlation between catchment size and the amount of genetic diversity of roach within the rivers. However, the degree to which diversity differs among the rivers is a function of historical diversity levels, founding population size and current carrying capacity and genealogical turnover. Before we can assess differentiation, and identify any causative agents, one must determine whether the differently sized and hydraulically distinct river systems harbor significantly different levels of genetic variation. One would expect the Thames population, with the river's greater capacity for harboring larger census populations, and with its increased hydrographical complexity, to possess greater levels of genetic diversity than would be found within the Stour. If migration, dispersal or anthropogenic translocation between these two rivers has been minimal since the last time they were connected via fluvial conduits (ca. 7500 years ago via the Channel River system), then one would expect that the greater influence of drift over mutation during the early years of colonization may lead to a significant divergence in allele frequencies between the two drainages. If so, such data would promote the hypothesis that physically distinct drainages, rather than multi-drainage assemblages, represent minimal base levels of conservation and preservation for management unit designation in riverine ecosystems.

If the Stour and Thames populations match significantly on all scores of diversity indices, this would provide some evidence that the two rivers share such a high degree of genetic diversity that they *could* be considered as a single, inter-breeding genetic entity (maintained through current gene flow, if equilibrium is assumed). However, as previously noted, high genetic diversity in contemporary populations may not be conducive to analyses of genetic differentiation for some categorical instances (e.g. in the number of alleles, or allelic richness, for instance), given the short geological time span since fish in both rivers would have last shared a common breeding pool of individuals. Therefore, if, on the other hand, divergence has occurred since physical separation, then mutation at microsatellite loci - which is random with respect to direction and to which loci actually transmutate - should not support a significant, predictive relationship between private allele frequencies between the Stour and the Thames. Thus, significant deviation from linear expectations for private alleles would be consistent with the expectations of high

ancestral genetic diversity, but recent post-Pleistocene separation allowing for some divergence, but not with a scenario in which the bidirectional flow of genetic material is an ongoing source of variation in these geographically proximate rivers[34].

4.3.2.2 Meaningful Levels of Genetic Structuring Between the Stour and Thames – Assessment of Equilibrium

The central contention of this section is that roach populations have attained mutation-drift equilibrium in both the Thames and the Stour, retaining ancestral levels of variation, but having census populations large enough that numerous new mutations have occurred in both populations (the probability of spontaneous mutation is a product of the mutation rate and the population size, $2N_e\mu$). Values of F_{ST} are expected to be low (< 0.05), although standardized estimators and Jost's D (Jost 2008) are expected to be > F_{ST}. Due to their levels of variation and historically high census sizes, both the Thames and the Stour should have some immunity to periodic, localized crashes in numbers, and thus retain a signal of 'mutation-drift equilibrium'. Additionally, the effect of mutation on levels of differentiation is dependent upon mutation rate, population size and time since isolation. Jost's D is demographically independent and so describes the extent of differentiation that may be ascribed to mutational effect alone in driving divergence between the Thames and the Stour. Given the expectation that the allelic state of some loci will be unique in each river, one would expect a significant proportion of divergence to be ascribed to the differential effect of microsatellite evolution in the face of weak levels of genetic drift.

Further, the issue of migration-drift equilibrium will be explored. If there has been little to negligible migration between the rivers' populations, then any divergence is the result of mutation and drift acting at an equilibrium rate describing the actual extent of genetic differentiation only when the populations' evolutionary dynamics no longer carry the signal of past gene flow and connectivity. If large populations fragment, then the prime determinant of differentiation – genetic drift – will be weak, and the number of generations needed to reach equilibrium levels of differentiation will be considerable. Most traditional measures of population differentiation, standardized or not, assume equilibrium conditions and may well underestimate the true level of isolation between populations (Allendorf and Luikart 2006). To rely on such estimates, one should test whether this null hypothesis is true, or whether populations better fit a purely-drift model, in which the degree to which the two populations are truly differentiated is probably

[34] Close, that is, in terms of distance from river mouth to river mouth.

underestimated (and therefore levels of current inter-river migration are overestimated) and the equilibrium level of differentiation is purely a function of population size (assuming little influence of mutation). If migration-drift equilibrium is found to be present then this may result from isolated populations sharing no recent migrants, but retaining a past signal of gene flow, or from recurrent inter-drainage gene flow. One may test for inter-population migration using non-equilibrium based tests to differentiate between these competing hypotheses to uncover a signal of a recent exchange of individuals.

4.3.3 Statistical Analysis.

4.3.3.1 Analysis of Microsatellite Diversity

All indices of genetic variation for the pooled Thames and Stour individuals were conducted in the same suite of population genetic software as was used to appraise the global levels for each of the ten study loci (section 4.2.4 above). Following this initial survey of diversity, linear regression analysis was carried out in the software package Mathematica 8.0 (Wolfram Research, Inc., 2010). The form of linear regression adopted in this comparison followed a strict linear fitted model[35], in which values of x (in this case, values of diversity from the Stour, which one would predict to be less variable than those from the Thames) are tested for significance in predicting the values of y (Thames diversity indices). In other words, by minimising the sum of the squares of the vertical distances between directly comparable points (e.g. the number of alleles at locus Ca12 for the Stour and the number of Ca12 alleles for the Thames, repeat for all other loci), the data are best fitted to a straight line (given by the equation: $y = a+bx$, where x = Stour, y = Thames, b = the slope of the line and a = the value of y when $x = 0$). A perfectly linear, predictive relationship would result in a perfectly straight line through all points when displayed on a graph.

In addition to traditional measures of genetic diversity, single point estimates of the effective population size of each of the respective 'populations' were determined from the pooled sample of individuals from within both rivers. Various point estimates of N_e have been developed over the years, each predicated on a set of assumptions and an inferential

[35] A reasonable assumption given the high probability that much of the genetic diversity shared between the two rivers is sourced from the same post-glacial colonisation route.

methodology particular to that method. A linkage disequilibrium (LD) analysis was carried out firstly in N_eEstimator version 1.3 (Peel et al. 2004), and then in LDN_e (Waples & Do 2008), a program in which a correction is applied to a downward bias in N_e estimation when the LD method is applied to a number of sampled individuals that is fewer than the actual number of breeding individuals (Waples 2006). Alleles with frequencies < 0.05 were discarded from the analysis, and estimates of N_e were calculated with 95% confidence intervals calculated from non-parametric bootstrapping over all loci.

4.3.3.2 Analysis of Genetic Structuring

Previous genetic studies of roach, in which nuclear genetic markers were used to measure equilibrium levels of genetic differentiation among populations, uncovered significantly high levels of intra-population genetic diversity for allozyme, RFLP and microsatellite marker loci. For all markers and all studies the greatest proportion of the variance of allele frequencies was overwhelmingly within populations as opposed to among populations. Most studies used the canonical metric of differentiation, Wright's F_{ST}, describing hierarchical levels of genetic structuring within and amongst populations that are assumed to be at mutation-migration-drift equilibrium.

Latterly, population genetic studies have used various estimators of F_{ST} (such as G_{ST}, θ and σ_{ST}) that have at their core a reliance upon assessing levels of genetic divergence amongst populations by relating the amount of genic diversity (expected heterozygosity) within populations to that found in the total population (Meirmans & Hedrick 2011). However, these estimators of genetic divergence underestimate the true level of genetic disparity among populations when levels of intra-population heterozygosity are high (Jost 2008). For highly variable loci, the maximum level of differentiation is not necessarily given by 1, instead such measures are influenced by high levels of genic diversity within populations, such that there is often a negative correlation between the amount of within-population genetic diversity and the maximum attainable value of the estimator of F_{ST} amongst populations (Meirmans & Hedrick 2011; Heller & Siegismund 2009). Standardization, whereby the estimator of F_{ST} is given as a function of its maximal possible value, has been put forward as a solution to the issue of highly genetically diverse populations in population genetic studies. Standardized values of G_{ST}, in which sampling from within a metapopulation of k sub-populations takes into account some of the uncertainty of not sampling all populations and/or of taking multiple samplings from

within a single population (G'_{ST}, Hedrick 2005b), θ (θ', Meirmans & Hedrick 2011) and φ_{ST} ($\varphi'_{ST,}$ Meirmans 2006) have all been developed in which the maximal value that can be expected when allelic differentiation over all populations is absolute is 1. Jost (2008) incepted a new measure of differentiation, D, in response to criticisms of F_{ST}-estimators based upon levels of heterozygosity because: i) H_S puts a limit on the amount of possible divergence; and ii) the unsuitability of H_S as a base for a divergence metric because of its non-linearity with actual levels of genetic diversity (Meirmans & Hedrick 2011). Jost's D is instead based on the effective number of alleles found within and among populations, which does scale linearly with diversity. Whilst based on the number of effective alleles, D can be related to heterozygosity by the second equation in Meirmans & Hedrick (2011) (Equation 11 in Jost (2008)).

The level of genetic divergence between the Thames and the Stour was initialy assessed by the application of Fisher's exact test of population differentiation, as implemented in Genepop version 4.0 (Raymond and Rousset 1995b; Rousset 2008). The software uses the MCMC sampling procedure (100000 dememorizations steps (analogous to 'burn-in'), with 1000 batches and 5000 iterations per batch) and the log likelihood ratio statistic, G (Guo and Thompson 1992), to detect significant deviations from a random distribution of genotypic frequencies between the two rivers (Goudet et al. 1996). Axiomatic estimates of population structuring - F_{ST}, R_{ST} - were calculated in the packages FSTAT version 2.9.3.2 (Goudet 1995) and SPAGeDi version 1.3 (Hardy & Vekemans 2002). Statistical significance for all estimates of population sub-structuring in FSTAT was determined by permuting the dataset 10000 times, whilst in SPAGeDi standard error estimates for global values were determined using the jack-knifing procedure of Sokal & Rohlf (1995) over all loci. Within FSTAT, Weir and Cockerham's (1984) F_{ST}-estimator, θ, was calculated for each locus (over both populations and all loci). 1000 permutations of the dataset allowed the construction of 95% confidence intervals. θ assumes that populations are at migration-mutation-drift equilibrium, a situation that may not be commonplace in nature (although this simple model approximates real data with surprising consistency)[36], especially if the markers employed do not follow the IAM model of evolution.

[36] When interpreting genetic divergence as a consequence of gene flow, or lack thereof, one assumes populations are at migration-drift equilibrium; when one assumes populations have recently diverged then genetic drift (the rate of fixation of alleles) becomes the causal locus for divergence (Hardy et al. 2003).

Bias-corrected estimates of both H_S & H_T (strictly only necessary when the number of sampled individuals is small) can be used to estimate the new class of standardized F_{ST} estimators, as well as Nei's D and Jost's D. Jost's D, however, takes longer to reach equilibrium than the standardized estimators of F_{ST}, particularly when rates of mutation are low, implying that D is more greatly affected by the mutation rate than the other estimates (Ryman & Leimar 2009). In the absence of migration, the fact that all new variation is derived from mutation best fits the expectation of the D statistic, which is itself independent of population size (Jost 2009). However, D is an unsuitable metric for highly connected populations experiencing high rates of gene flow and/or low mutation rates (Meirmans & Hedrick 2011). Whilst Jost's D is seemingly the most appropriate measure of divergence for demographically isolated populations, the effect of demography and genetic drift are neglected (the loss of heterozygosity is inversely related to population size). Therefore, traditional standardized measures of differentiation are also included for comparison.

G''_{ST}, a bias-reducing correction of G'_{ST} (Meirmans & Hedrick 2011), was calculated for both the actual and simulated datasets based on allele frequencies using the software Genodive version 2.0b21 (Meirmans & van Tienderen 2004). The same software was also used to calculate Jost's D (which already includes a correction term for a small number of compared populations) and θ', the calculation of which is dependent upon the maximum value of θ being derived from a re-coding of the dataset such that each sampled population possesses only those alleles that are unique to that population (Meirmans 2006). Standard errors, to quantify the variance in differentiation estimation attributable per locus and to mean values over all loci, were calculated by employing a jackknifing procedure inclusive within the software package

Furthermore, one applied assignment tests upon all 1001 roach individuals to assess the probability that each may be assigned to its population of origin based upon the differential allele frequencies in the two reference river populations. The analysis also assigned groups of individuals, i.e. each of the sampled 'populations', to each of the two rivers. If there is genetic differentiation between the Stour and the Thames, then the success of assigning individuals and populations to their respective sampled rivers should be high. This analysis was undertaken in GeneClass 2.0 **(**Piry et al. 2004**).** Geneclass possesses three inferential methodologies: a Bayesian approach (Rannala & Mountain 1997); a frequency-based method (Paetkau et al. 1995); and a genetic distance method (Cornuet et al. 1999). These methods are listed in decreasing order of accuracy in

simulation studies based upon 10 populations, 10 independent loci, 30 individuals per population, and an average F_{ST} of 0.1 (Cornuet et al. 1999). However, with these conditions, assignments generally achieve 100% success rates (Primmer et al. 2000). All methods adopted the invocation that each individual under analysis is removed from the reference datasets. For the distance-based method, two distances were analyzed: the chord distance of Cavalli-Sforza & Edward (1967) and Nei's minimum genetic distance (D_{NEI}) (Nei 1973).

4.3.3.3 Testing the Influence of Microsatellite Mutational Model on Estimates of Genetic Structure

The stepwise, bidirectional evolution of microsatellite markers, suggests that in some cases, alleles that are identical in state (IBS) in two populations may not be identical by descent (IBD), i.e. they have different evolutionary histories, and the populations are actually more genetically and evolutionarily divergent than a naïve recording of allelic state frequencies would suggest. Slatkin (1985) incorporated the ladder-like evolution of microsatellite loci into the hierarchical model of *F*-statistics. R_{ST} is analogous to F_{ST}, such that R_{ST} describes the correlation of allele sizes between loci within populations (Slatkin 1995; Hardy et al. 2003). Slatkin's method (1995) assumes that the expected squared differences among allele size classes closely approximate the linear function of the time to coalescence of any paired alleles at a locus. The method records a 'mutational memory', such that the number of stepwise differences between two alleles increases in time since common-ancestry (Hardy et al. 2003).

However, not all microsatellites faithfully adhere to the SMM model (e.g. Estoup and Angers 1998; Ellegren 2000b), and thus R_{ST} estimates may be biased due to the underlying assumption of the mutational process. If the SMM model of microsatellite evolution is upheld, F_{ST} and R_{ST} estimates that are similar for an identical set of data indicates that the genetic differentiation among two or more populations may be attributed to drift. If, however, R_{ST} is > F_{ST}, then one may attribute the differences in calculated divergence to the influence of the microsatellite mutational model assumed, with the result being that F_{ST} would be underestimated if IAM were to be adopted in place of the SMM. To determine whether the mutational dynamics of microsatellite evolution influences the estimation of population divergence and genetic structuring, Hardy's allele size randomization test (Hardy et al. 2003), in which allele sizes are randomly permuted among allelic states, was implemented in SPAGeDi version 1.2 (Hardy and Vekemans 2002). The

allelic size data were permuted the maximum 20000 times to determine the existence of a variant expectation to the null expectation of a zero effect of allelic motif size on diversity and differentiation estimates, in relation to either the rate of migration among the two rivers (facilitated naturally or anthropogenically) or the inverse of the time since isolation ($1/t$). The analysis formulated estimates of uncorrected R_{ST} and R_{ST} corrected for size-permutations (pR_{ST}), the latter having an expected value lower than the unpermuted R_{ST} estimate if the SMM contributes greatly to the non-random association of allelic state and microsatellite repeat motif number.

4.3.3.4 Deviation from Mutation-Migration-Drift Equilibrium

The degree to which two demographically isolated, panmictic populations are genetically divergent is dependent upon four factors: shared population histories, the mutation rate, the population size (the power accorded to genetic drift) and the time (number of generations) that has elapsed since the two populations last shared a common breeding pool. If, however, one or more of the populations has suffered a recent population bottleneck, or the time that has elapsed has not been long enough to erode away the signals of a founding bottleneck, then the amount of divergence exhibited between the two populations should be greater than that expected under mutation-drift equilibrium. The software BOTTLENECK (Piry et al. 1999) was implemented to test whether the assumption of mutation-drift equilibrium holds for each of the pooled datasets of the Stour and the Thames, respectively. Under a bottlenecking scenario, heterozygosity is lost at a rate considerably less than the number of unique alleles, with rare alleles being more likely to be expunged from a population than more common ones. Therefore, the observed heterozygosity will be greater than the expected heterozygosity of the same population under a model of mutation-drift equilibrium ($H_o > H_e$) (Luikart et al. 1998), at least within a period of time that approximates $0.2 – 4N_e$ generations (Luikart & Cornuet 1998). BOTTLENECK uses coalescent simulations to derive, from the observed number of alleles, distributions of heterozygosities under the null hypothesis of equilibrium conditions and under non-equilibirum conditions, for three different models of evolution (IAM, SMM and a two-phase model (TPM) in which most mutations follow the SMM, but a pre-determined proportion (k) follow the IAM (Di Rienzo et al. 1994)). All three models were implemented in BOTTLENECK, within which the Wilcoxon-signed ranks test was performed to determine the significance of any skew from a neutral H_o:H_e ratio. This test is considered the most statistically powerful available in BOTTLENECK for any analysis in which there are less than 20 marker loci (Piry et al. 1999). For illustration, the shape of

the frequency spectrum of allele numbers can qualitatively indicate a deviation from equilibrium, with deviations from a smooth L-shaped curve indicating such a process (Luikart et al. 1998).

Finally, one tested whether the two river populations share a signal of recent or historical gene flow, or whether they meet the expectation of genetic diversity distributed under a model of isolation and genetic drift alone after fragmentation from a common source. To test the latter, the program 2mod version 2.0 (Ciofi et al. 1999) utilizes coalescent-simulations to see if the data best fits an equilibrium migration-drift model or whether it best fits a model of genetic drift only. By employing an MCMC approach, the program compares the likelihoods of each model given the allele frequency data of each river. The program was run for each model over 100000 iterations (of which the first 10% were discarded). The numbers of iterations that support each model were tallied and the most likely model is the one with the greatest tally.

4.3.4 Results

4.3.4.1 Levels of Genetic Diversity within the Thames and the Stour

Tables 4.7 and 4.8 show estimates of genetic diversity for all sampled locations within the Thames and Stour rivers, respectively. Despite some locales providing few or no successful PCR amplifications for certain loci (hence the variable number of genotyped individuals for each locus), overall levels of genetic diversity seem remarkably similar at first glance. The mean numbers of alleles (18 versus 17 in the Stour and in the Thames, respectively), number of private alleles (2.7 as opposed to 2.5), number of effective alleles (5.450 versus 5.735) and allelic richness (17.414 vs. 17.159) are all found within a margin of 8% of the higher figure. Moreover, the allele frequencies for both the Stour and the Thames are broadly co-distributed (see Appendix B).

Of the 129 possible locus-location comparisons in the Thames (Table 4.7), 10 were statistically significant for deviations from HWE after the application of the Bonferroni correction, whereas 43 are significant before applying the correction.

Table 4.7: Indices of genetic diversity for 13 sampled sites within the Thames. N_S = sample size; N_A = number of unique alleles per location; A_R = allelic richness; H_E = expected heterozygosity; H_O = observed heterozygosity; F_{IS} = population inbreeding coefficient.

Locus	Sample Site												
	MWP	DL	OW	CI	Te	Wh	Dor	Day	Cu	Ey	No	Bu	Ro
Lid1													
N_S	32	64	48	61	57	23	31	31	40	32	23	26	21
N_A	8.000	10.00	10.00	8.000	9.000	10.00	7.000	9.000	9.000	10.00	6.000	6.000	6.000
A_R	6.450	7.881	7.910	6.174	6.156	8.888	5.810	8.206	6.321	8.439	5.857	5.192	5.429
H_E	0.702	0.813	0.743	0.657	0.676	0.776	0.729	0.811	0.649	0.741	0.644	0.655	0.668
H_O	0.563	0.609	0.583	0.541	0.632	0.652	0.613	0.484	0.750	0.750	0.652	0.731	0.619
F_{IS}*	0.201	0.252	0.216	0.177	0.067	0.162	0.161	0.407	-0.157	-0.012	-0.012	-0.119	0.075
Lco4													
N_S	33	61	48	61	56	22	31	31	40	32	22	25	22
N_A	5.000	5.000	5.000	6.000	7.000	4.000	6.000	6.000	6.000	6.000	6.000	5.000	6.000
A_R	4.357	4.300	4.570	4.916	6.542	3.991	5.891	5.671	5.806	4.922	5.763	4.796	5.722
H_E	0.449	0.621	0.691	0.646	0.778	0.673	0.769	0.707	0.743	0.613	0.755	0.643	0.714
H_O	0.364	0.656	0.646	0.508	0.714	0.500	0.710	0.548	0.600	0.531	0.500	0.520	0.773
F_{IS}	0.192	-0.057	0.066	0.215	0.083	0.262	0.079	0.227	0.195	0.135	0.343	0.195	-0.085
Rru3													
N_S	32	62	48	57	51	17	21	33	38	33	24	23	22
N_A	5.000	5.000	4.000	7.000	5.000	5.000	4.000	5.000	4.000	5.000	7.000	5.000	5.000
A_R	4.892	4.693	3.984	5.534	4.028	5.000	4.000	4.508	3.987	4.489	6.315	4.721	4.989
H_E	0.596	0.576	0.567	0.644	0.492	0.652	0.661	0.578	0.587	0.528	0.665	0.471	0.661
H_O	0.500	0.290	0.479	0.351	0.451	0.588	0.476	0.333	0.500	0.364	0.500	0.391	0.636
F_{IS}	0.163	0.498	0.156	0.458	0.084	0.101	0.284	0.427	0.150	0.315	0.252	0.173	0.038
Ca1													
N_S	25	60	49	55	53	23	30	32	34	31	25	24	21
N_A	8.000	12.00	13.00	14.00	10.00	9.000	6.000	8.000	9.000	10.00	9.000	10.00	7.000
A_R	7.453	8.625	8.867	9.407	7.200	7.829	5.371	6.820	7.266	8.147	8.066	9.034	6.554
H_E	0.796	0.822	0.792	0.839	0.783	0.743	0.711	0.796	0.764	0.798	0.813	0.845	0.753
H_O	0.760	0.683	0.818	0.727	0.698	0.652	0.600	0.719	0.882	0.645	0.440	0.875	0.619
F_{IS}	0.046	0.170	-0.034	0.134	0.110	0.125	0.158	0.098	-0.158	0.194	0.464	-0.036	0.181
Ca3													
N_S	31	58	46	60	38	23	30	29	37	32	24	18	21
N_A	16.00	17.00	17.00	19.00	22.00	15.00	18.00	16.00	16.00	19.00	16.00	18.00	16.00
A_R	13.91	14.39	13.49	14.27	16.06	12.98	15.17	13.48	13.40	15.96	14.22	15.17	14.74
H_E	0.929	0.936	0.928	0.931	0.944	0.903	0.940	0.911	0.926	0.941	0.925	0.934	0.931
H_O	0.839	0.931	0.870	0.917	0.947	0.870	0.700	0.724	0.730	0.750	0.833	0.800	0.667
F_{IS}	0.099	0.005	0.064	0.015	-0.003	0.038	0.258	0.208	0.215	0.206	0.101	0.146	0.289

* Each estimate of the population inbreeding coefficient (F_{IS}) is considered significantly different from a null expectation of HWE if highlighted yellow and underlined (significant regardless of Bonferroni correction) or shaded grey (only significant before Bonferroni correction).

Table 4.7: Continued.

Locus	Sample Site												
	MWP	DL	Wi	CI	Te	Wh	Dor	Day	Cu	Ey	No	Bu	Ro
Ca12													
N_S	29	65	49	62	60	22	33	33	34	33	25	25	22
N_A	17.00	21.00	23.00	23.00	19.00	16.00	15.00	15.00	17.00	13.00	13.00	16.00	15.00
A_R	12.85	13.25	13.99	13.38	13.08	14.06	11.85	11.26	13.08	10.33	10.46	13.01	13.16
H_E	0.869	0.805	0.854	0.830	0.854	0.865	0.829	0.785	0.864	0.732	0.676	0.793	0.784
H_O	0.897	0.800	0.857	0.726	0.817	0.818	0.727	0.636	0.794	0.697	0.680	0.720	0.682
F_{IS}	-0.032	0.006	-0.003	0.126	0.044	0.055	0.124	0.192	0.082	0.048	-0.006	0.093	0.133
Cy3													
N_S	28	65	45	58	57	22	32	31	39	33	22	26	22
N_A	17.00	18.00	16.00	20.00	22.00	15.00	15.00	13.00	18.00	14.00	15.00	15.00	13.00
A_R	12.43	11.10	11.72	12.49	13.11	13.34	11.08	9.767	12.66	10.76	12.84	11.99	11.29
H_E	0.755	0.746	0.811	0.870	0.798	0.818	0.706	0.700	0.785	0.720	0.808	0.739	0.697
H_O	0.679	0.600	0.667	0.724	0.772	0.682	0.656	0.645	0.692	0.727	0.955	0.654	0.545
F_{IS}	0.103	0.197	0.180	0.169	0.033	0.170	0.072	0.079	0.119	-0.011	-0.187	0.117	0.221
Cy48													
N_S	33	51	30	56	58	23	32	33	38	32	26	0	20
N_A	15.00	16.00	12.00	14.00	14.00	11.00	13.00	13.00	11.00	12.00	13.00	NA**	11.00
A_R	12.40	11.14	10.25	10.83	10.51	9.631	11.21	11.04	9.400	10.02	11.88	NA	10.04
H_E	0.912	0.889	0.864	0.884	0.885	0.850	0.909	0.896	0.840	0.889	0.914	NA	0.888
H_O	0.939	0.922	0.700	0.857	0.759	0.739	0.844	0.879	0.816	0.625	0.769	NA	0.900
F_{IS}	-0.030	-0.037	0.193	0.031	0.144	0.133	0.073	0.019	0.029	0.300	0.161	NA	-0.013
Lc27													
N_S	33	59	49	63	60	23	33	33	40	32	26	26	20
N_A	3.000	5.000	4.000	3.000	4.000	3.000	4.000	3.000	3.000	4.000	4.000	3.000	4.000
A_R	2.746	3.568	3.355	2.263	2.955	2.739	3.746	2.514	2.892	3.308	3.304	2.952	3.848
H_E	0.197	0.311	0.243	0.225	0.227	0.300	0.426	0.265	0.287	0.255	0.246	0.247	0.419
H_O	0.152	0.271	0.245	0.222	0.200	0.261	0.455	0.242	0.275	0.219	0.269	0.231	0.350
F_{IS}	0.234	0.130	-0.008	0.014	0.121	0.134	-0.069	0.087	0.042	0.146	-0.097	0.065	0.169
Lc290													
N_S	33	62	49	63	57	18	32	33	38	31	25	27	20
N_A	10.00	10.00	8.000	8.000	9.000	7.000	8.000	9.000	9.000	10.00	7.000	10.00	10.00
A_R	8.448	7.514	6.592	6.743	7.105	6.887	7.213	7.096	8.138	8.269	6.621	9.226	9.360
H_E	0.795	0.801	0.746	0.791	0.756	0.797	0.821	0.723	0.786	0.808	0.798	0.869	0.814
H_O	0.636	0.677	0.694	0.730	0.754	0.667	0.594	0.485	0.553	0.710	0.680	0.630	0.800
F_{IS}	0.202	0.155	0.070	0.077	0.002	0.167	0.280	0.333	0.299	0.124	0.150	0.279	0.018
All													
N_S	30.90	60.70	45.60	59.60	54.70	21.60	30.50	31.90	37.80	32.10	24.20	22.70	21.10
N_A	10.40	11.90	11.20	12.20	12.10	9.500	9.600	9.700	10.20	10.30	9.600	9.778	9.300
A_R	8.592	8.645	8.473	8.600	8.675	8.534	8.133	8.036	8.223	8.463	8.533	8.454	8.475
H_E	0.700	0.732	0.724	0.732	0.719	0.738	0.750	0.717	0.723	0.703	0.724	0.688	0.733
H_O	0.633	0.644	0.656	0.630	0.674	0.643	0.637	0.570	0.659	0.602	0.628	0.617	0.659
F_{IS}	0.098	0.121	0.095	0.140	0.063	0.131	0.152	0.208	0.089	0.145	0.136	0.106	0.103

** NA indicates that not enough individuals were typed for this locus at this location to be included in the analysis.

Table 4.8: Indices of genetic diversity for 9 sampled sites within the Stour. N_S = sample number; N_A = number of unique alleles per location; A_R = allelic richness; H_E = expected heterozygosity; H_O = observed heterozygosity; F_{IS} = population inbreeding coefficient.

Locus	Sample Site								
	Br	**DM**	**St**	**AB**	**Sh**	**MM**	**RC**	**SbC**	**Th**
Lid1									
N_S	82	40	71	54	59	66	18	42	20
N_A	9.000	7.000	11.000	17.000	6.000	7.000	12.000	7.000	7.000
A_R	5.371	4.628	7.204	8.080	4.845	4.829	10.035	5.642	6.159
H_E	0.626	0.604	0.800	0.796	0.638	0.633	0.850	0.743	0.597
H_O	0.577	0.559	0.547	0.662	0.656	0.576	0.882	0.583	0.550
F_{IS*}	0.079	0.076	0.318	0.170	-0.028	0.090	-0.039	0.217	0.081
Lco4									
N_S	86	39	82	58	64	72	19	43	20
N_A	5.000	3.000	5.000	10.000	3.000	5.000	6.000	5.000	6.000
A_R	3.657	2.896	3.724	5.541	2.833	3.666	5.601	4.177	5.143
H_E	0.586	0.525	0.565	0.728	0.551	0.590	0.740	0.606	0.704
H_O	0.549	0.529	0.466	0.522	0.588	0.516	0.529	0.270	0.700
F_{IS}	0.063	-0.008	0.177	0.285	-0.069	0.126	0.291	0.558	0.006
Rru3									
N_S	84	36	71	56	21	66	19	40	19
N_A	6.000	6.000	7.000	13.000	NA**	8.000	7.000	5.000	6.000
A_R	4.659	5.227	3.998	7.906	NA	4.933	6.402	4.320	5.206
H_E	0.691	0.790	0.759	0.806	NA	0.756	0.806	0.698	0.629
H_O	0.444	0.469	0.566	0.333	NA	0.557	0.706	0.639	0.579
F_{IS}	0.359	0.410	0.256	0.588	NA	0.264	0.127	0.085	0.081
Ca1									
N_S	83	40	65	53	35	63	19	38	19
N_A	9.000	6.000	11.000	13.000	10.000	12.000	5.000	8.000	8.000
A_R	6.410	4.919	7.010	7.072	8.023	7.541	4.961	5.925	6.635
H_E	0.779	0.724	0.816	0.789	0.775	0.831	0.759	0.765	0.744
H_O	0.605	0.588	0.528	0.574	0.758	0.690	0.600	0.618	0.579
F_{IS}	0.224	0.190	0.355	0.274	0.023	0.171	0.215	0.195	0.227
Ca3									
N_S	69	23	67	51	33	48	12	33	18
N_A	20.000	15.000	23.000	22.000	17.000	19.000	13.000	18.000	14.000
A_R	11.712	11.588	12.749	13.681	12.687	12.945	12.347	11.545	11.704
H_E	0.919	0.919	0.928	0.940	0.933	0.938	0.939	0.905	0.914
H_O	0.754	0.783	0.873	0.721	0.800	0.721	0.929	0.750	0.778
F_{IS}	0.181	0.151	0.060	0.234	0.145	0.234	0.012	0.174	0.153

* Each estimate of the population inbreeding coefficient (F_{IS}) is considered significantly different from a null expectation of HWE if highlighted yellow and underlined (significant regardless of Bonferroni correction) or shaded grey (only significant before Bonferroni correction).
** NA indicates that not enough individuals were typed for this locus at this location to be included in the analysis.

The majority of cases deviate significantly from 'random mating' by exhibiting a deficit of heterozygotes (global mean observed heterozygosity ranges from 0.570 – 0.659 across the 13 sample sites; F_{IS} spans 0.063 – 0.208; $p < 0.0001$ in 9 of the 13 sites). There is no general pattern within particular Thames locations with regards to deviations from expected heterozygosity, or a systematic association of heterozygote deficit with a particular locus.

Table 4.8: Continued.

Locus	Sample Site								
	Br	**DM**	**St**	**AB**	**Sh**	**MM**	**RC**	**SbC**	**Th**
Ca12									
N_S	81	40	71	54	59	66	18	42	20
N_A	21.000	17.000	17.000	23.000	14.000	16.000	8.000	16.000	8.000
A_R	10.383	11.473	10.956	10.921	8.767	9.375	7.016	10.661	6.337
H_E	0.860	0.907	0.886	0.894	0.846	0.803	0.761	0.856	0.704
H_O	0.785	0.824	0.807	0.700	0.824	0.767	0.588	0.774	0.500
F_{IS}	0.088	0.094	0.090	0.218	0.027	0.045	0.233	0.097	0.295
CypG3									
N_S	78	30	74	55	44	58	16	17	20
N_A	17.000	11.000	16.000	18.000	10.000	16.000	9.000	9.000	9.000
A_R	7.111	7.157	7.259	8.148	6.589	8.336	7.500	7.626	6.292
H_E	0.582	0.545	0.583	0.706	0.542	0.692	0.778	0.607	0.440
H_O	0.579	0.552	0.569	0.530	0.563	0.588	0.375	0.500	0.450
F_{IS}	0.005	-0.012	0.024	0.251	-0.038	0.151	0.526	0.181	-0.024
CypG48									
N_S	73	38	20	51	38	62	0	39	14
N_A	16.000	11.000	13.000	19.000	10.000	14.000	NA	11.000	9.000
A_R	9.684	8.795	10.405	11.278	8.208	9.486	NA	8.339	8.425
H_E	0.884	0.885	0.874	0.910	0.861	0.891	NA	0.862	0.847
H_O	0.861	0.912	0.850	0.846	0.765	0.774	NA	0.778	0.786
F_{IS}	0.026	-0.031	0.029	0.071	0.113	0.133	NA	0.099	0.074
Lc27									
N_S	82	40	83	57	63	76	19	39	19
N_A	5.000	2.000	5.000	5.000	4.000	3.000	2.000	2.000	4.000
A_R	2.290	1.991	2.836	2.595	2.442	2.199	1.998	1.960	3.853
H_E	0.135	0.255	0.208	0.232	0.141	0.136	0.226	0.187	0.477
H_O	0.141	0.235	0.190	0.214	0.147	0.127	0.250	0.206	0.474
F_{IS}	-0.048	0.077	0.091	0.077	-0.041	0.067	-0.111	-0.100	0.006
Lc290									
N_S	86	37	73	52	56	73	16	42	18
N_A	7.000	7.000	9.000	9.000	7.000	10.000	9.000	10.000	8.000
A_R	5.879	5.431	6.570	7.215	5.563	6.608	8.188	7.161	6.968
H_E	0.758	0.720	0.779	0.843	0.755	0.776	0.863	0.839	0.827
H_O	0.695	0.559	0.667	0.716	0.742	0.794	0.813	0.583	0.667
F_{IS}	0.084	0.226	0.145	0.151	0.018	-0.022	0.060	0.308	0.198
All									
N_S	80.4	36.3	68.6	54.4	47.6	65.5	15.7	36.8	18.7
N_A	11.500	8.500	11.700	14.900	9.000	11.000	7.889	9.100	7.900
A_R	6.716	6.411	7.271	8.244	6.662	6.992	7.116	6.736	6.672
H_E	0.682	0.687	0.720	0.764	0.672	0.705	0.747	0.707	0.688
H_O	0.599	0.601	0.606	0.582	0.649	0.611	0.630	0.570	0.606
F_{IS}	0.122	0.128	0.159	0.240	0.034	0.134	0.160	0.196	0.122

Of the 88 locus-location comparisons within the Stour (Table 4.8), 20 revealed significant deviations from HWE expectations regardless of Bonferroni correction, 22.73% of comparisons against 7.76% in the Thames. 54 locus-by-location comparisons are significantly different from zero prior to Bonferroni correction. Like the Thames, overall estimates of observed heterozygosity (0.570-0.649; F_{IS}: 0.034 – 0.240, $p < 0.001$ in 6 out of 9 sites) for the nine sample sites display a deficit of heterozygotes. Locus Rru3 shows highly significant deficits of heterozygotes in 5/8 populations in which it was genotyped.

The mean number of alleles per population was 10.444±1.05 in the Thames and 10.165±2.181 in the Stour, and the mean allelic richness over all populations was 8.449±0.198 and 6.980±0.511 for the Thames and Stour, respectively. The Thames populations' mean levels of gene diversity, observed heterozygosity and population-level inbreeding were also higher than that found in the Stour (H_E = 0.722±0.017 and 0.708±0.029; H_O = 0.635±0.028 and 0.606±0.022; and F_{IS} = 0.100±0.028 and 0.095±0.053, for the Thames and Stour respectively).

The allele frequencies for all ten loci, pooled across the two rivers, embrace the range of frequency patterning often found with microsatellites (e.g. irregular, bimodal, or trimodal distributions (Valdés et al. 1993)), consistent with the expectations of the stepwise mutation model of microsatellite evolution. Whilst the mean values tend to be similar among the Stour and Thames individuals, there are locus specific differences. For example, there exist a differential number of alleles, greater than one, between the two rivers for the loci: Rru3, Lc290, Ca1 and Ca12. Furthermore, the number of private alleles, whilst tallying to similar mean values, is distributed among different loci. Similarly, there is also variation found among loci for the values of effective allele number and allelic richness between the Thames and the Stour populations. However, overall levels of variation, in the numbers of alleles and in their general patterning within the two rivers, appear very similar (see Fig 4.5).

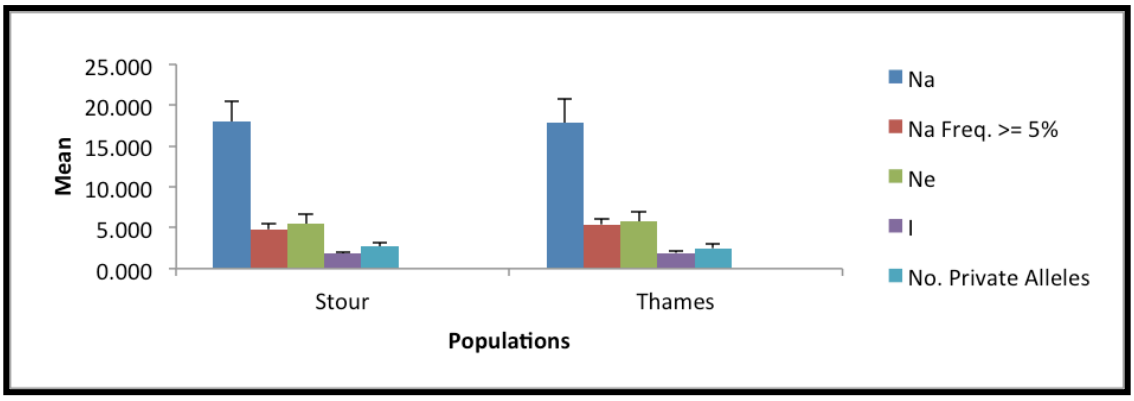


Figure 4.5: Graphical representations of mean values for a range of diversity indices. Na – number of alleles; Na > 5% = number of alleles with a frequency greater than 5%; Ne = number of effective alleles; I = Shannon's diversiy index. Bars represent the standard errors associated with the mean values for each of the diversity measures.

The results of the linear regression analysis are presented in Table 4.9 and represented graphically in Fig 4.6. Table 4.9 shows the estimate of both *x* and *a* (the value of *y* when *x* is zero), when the data is best fitted for a strictly linear, predictive relationship between the twin datasets of the Stour and the Thames. Also presented are the 95% confidence intervals for the values of *x* and *a* (top and bottom, respectively), the standard errors, and the t-statistics and *p*-values. Assuming a strict linear relationship, the diversity values of the Thames are directly predictable from the values of the Stour data for each diversity category (NA: $y = 1.1404$, $p < 0.001$; NEA: $y = 0.9483$, $p < 0.001$; A_r: $y = 1.0818$, $p < 0.001$; H_o: $y = 0.9255$, $p < 0.001$; H_e: $y = 0.8163$, $p < 0.001$; F_{IS}: $y = 0.6726$, $p = 0.0196$), with the exception of the number of private alleles (NPA: $y = 0.3257$, $p = 0.3976$).

Table 4.9: Results of the linear regression analyses. x and a refer to the parameters that describe a relationship between two variables (slope) (see text). For each diversity metric the top figure is x, whereas the bottom figure is a. 95% CI, standard error, the t-statistic and statistical significance is given for both values and for each diversity metric.

Diversity	x/a	95% CI		Standard Error	*t*-Statistic	*p*-value
		Lower	Upper			
NA	-2.7263	-6.2768	0.8241	1.5397	-1.7707	0.1146
	1.1404	0.9585	1.3223	0.0789	14.4569	5.12×10^{-7}
NEA	0.5664	-0.9724	2.1053	0.6673	0.8488	0.4207
	0.9483	0.7137	1.1828	0.1017	9.3238	1.42×10^{-5}
A_r	-1.6797	-5.2864	1.9270	1.5641	-1.0739	0.3142
	1.0818	0.8916	1.2720	0.0825	13.1158	1.09×10^{-6}
NPA	1.6207	-1.0236	4.2649	1.1467	1.4134	0.1953
	0.3257	-0.5147	1.1661	0.3644	0.8936	0.3976
H_o	0.0855	-0.0852	0.2561	0.0740	1.1552	0.2814
	0.9255	0.6508	1.2001	0.1191	7.7695	5.38×10^{-5}
H_e	0.1580	-0.0955	0.4116	0.1100	1.4374	0.1886
	0.8163	0.4761	1.1566	0.1476	5.5325	5.52×10^{-5}
F_{IS}	0.0341	-0.0626	0.1307	0.0419	0.8130	0.4397
	0.6726	0.1396	1.2056	0.2311	2.9101	0.0196

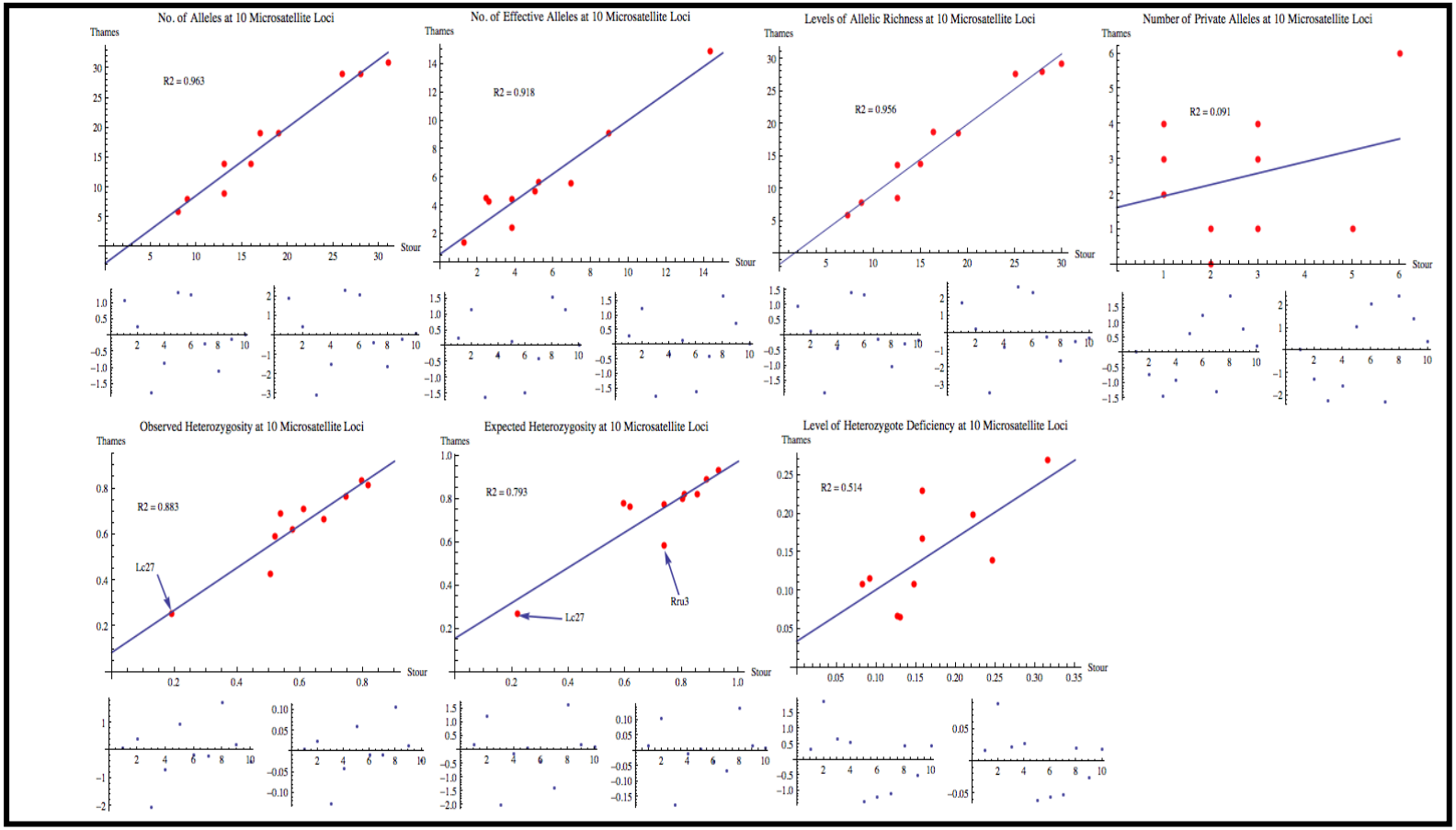


Figure 4.6: Linear regression plots for each of the seven genetic diversity measures listed in Table 4.8. Top row, left to right: number of alleles; number of effective alleles; allelic richness, number of private alleles. Bottom row, left to right: observed heterozygosity; expected heterozygosity; and heterozygote deficiency (F_{IS}). The smaller graphs under the main plots, show the distribution of the residuals for the Thames (left) and Stour (right), respectively.

Fig. 4.7 presents the results of a correlation analysis to determine the occurrence of predictive relationships between levels of genetic diversity averaged over all loci and the linearly ordered geographical sites at which roach were sampled, from downstream to upstream. For both rivers, the average number of alleles per site decreases with increasing upstream location, attaining statistical significance in the Thames (Pearson's $r = 0.6709$, 2-tailed $p = 0.0121$) (R^2 value shown for MNA in Fig 4.7 only), but not in the Stour (Pearson's $r = 0.4758$, 2-tailed $p = 0.1955$). No significant downstream-upstream trend is seen for allelic richness (Thames: Pearson's $r = 0.3492$, 2-tailed $p = 0.2422$; Stour: Pearson's $r = 0.0643$, 2-tailed $p = 0.8695$), gene diversity (Thames: Pearson's $r = 0.154$, 2-tailed $p = 0.6134$; Stour: Pearson's $r = 0.1168$, 2-tailed $p = 0.7647$), observed heterozygosity (Thames: Pearson's r = 0.2175, 2-tailed $p = 0.4753$; Stour: Pearson's $r = 0.0235$, 2-tailed $p = 0.9521$) or F_{IS} (Thames: Pearson's $r = 0.1545$, 2-tailed $p = 0.6143$; Stour: Pearson's $r = 0.0803$, 2-tailed $p = 0.8373$).

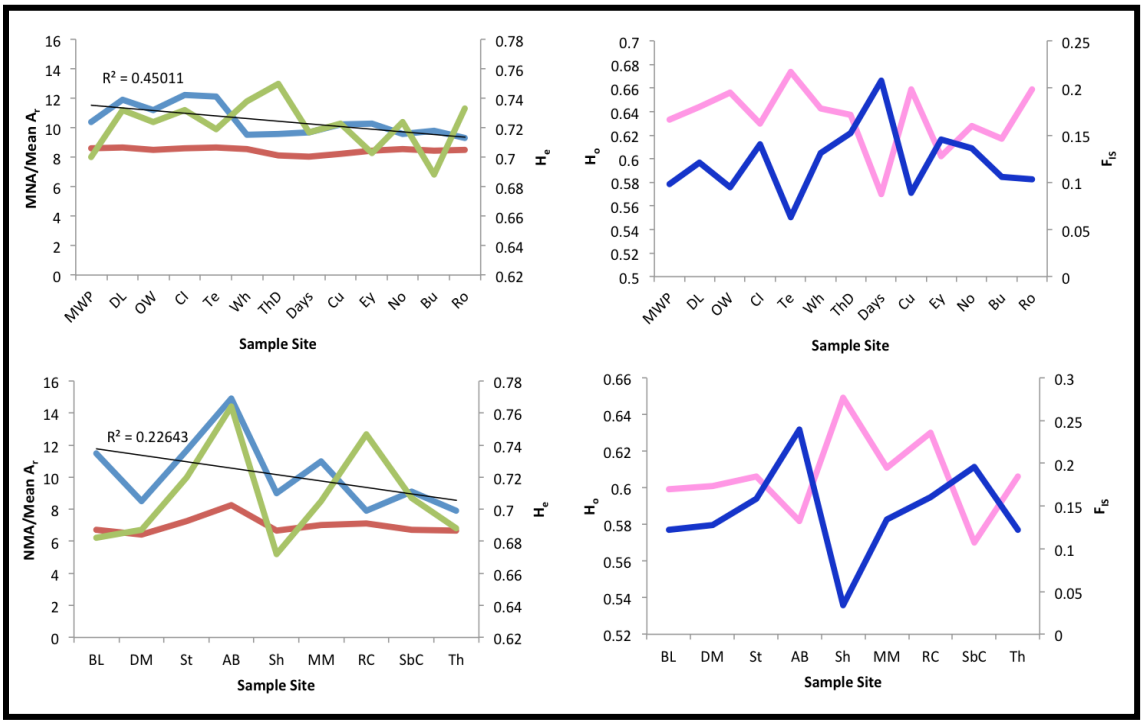


Figure 4.7: The relationship between mean number of alleles (MNA: light blue), allelic richness (A_r: red), gene diversity (H_e: green) (left) and observed heterozygosity (H_o: pink) and F_{IS} (navy blue) (right) with sample site (downstream to upstream (left to right)) for all loci combined. Thames (top); Stour (bottom).

Levels of allelic richness remain relatively stationary with increasing distance upstream in the Thames, with little perturbation, except for a slight dip between Dorchester and Days. In the Stour, however, allelic richness peaks at Anchor Bridge, declining immediately upstream before a smaller, secondary peak at Rat's Castle. The Anchor Bridge 'population' also experiences a large increase in the mean number of alleles as well as

gene diversity, with a concomitant decrease in observed heterozygosity commensurate with a reduction in the expected number of heterozygotes. The contiguous upstream 'population' at Shalford Weir shows a converse patterning. The sampled location of Rat's Castle – Stoke by Clare also shows a strong inverse relationship between observed heterozygosity and F_{IS}. The Dorchester-Days area of the Thames shows a marked reduction in observed heterozygosity relative to HWE expectations, in addition to an excess of homozygotes with a concomitant reduction in allelic diversity.

Fig 4.8 reveals the results of an analysis to determine whether the datasets from both rivers significantly differ in average values of diversity across the sampled sites. The smooth histograms (upper graph) plot the probability density function for each of the five diversity indices for both rivers, whereas the box-whisker plots (lower plot) graphically display the range (including 50% quartiles) of values found in each river averaged over all populations. Of the five comparisons, allelic richness (Mann-Whitney's U = 114, $p < 0.001$) and observed heterozygosity (Mann-Whitney's U = 94, $p = 0.0194$) were found to significantly deviate from the null expectation of no difference, with higher levels of both found in the Thames (Thames: mean $A_r = 8.447 \pm 0.198$; mean $H_o = 0.635 \pm 0.027$; Stour: mean $A_r = 6.980 \pm 0.542$; Mean $H_o = 0.606 \pm 0.023$). The remaining factors were found to overlap (see Fig 4.8) indicating no difference in mean levels of these parameters (MNA: Mann-Whitney's U = 75, $p = 0.2852$; H_e: Mann-Whitney's U = 0.125, $p = 0.1508$; F_{IS}: Mann-Whitney's U = 39, $p = 0.1816$).

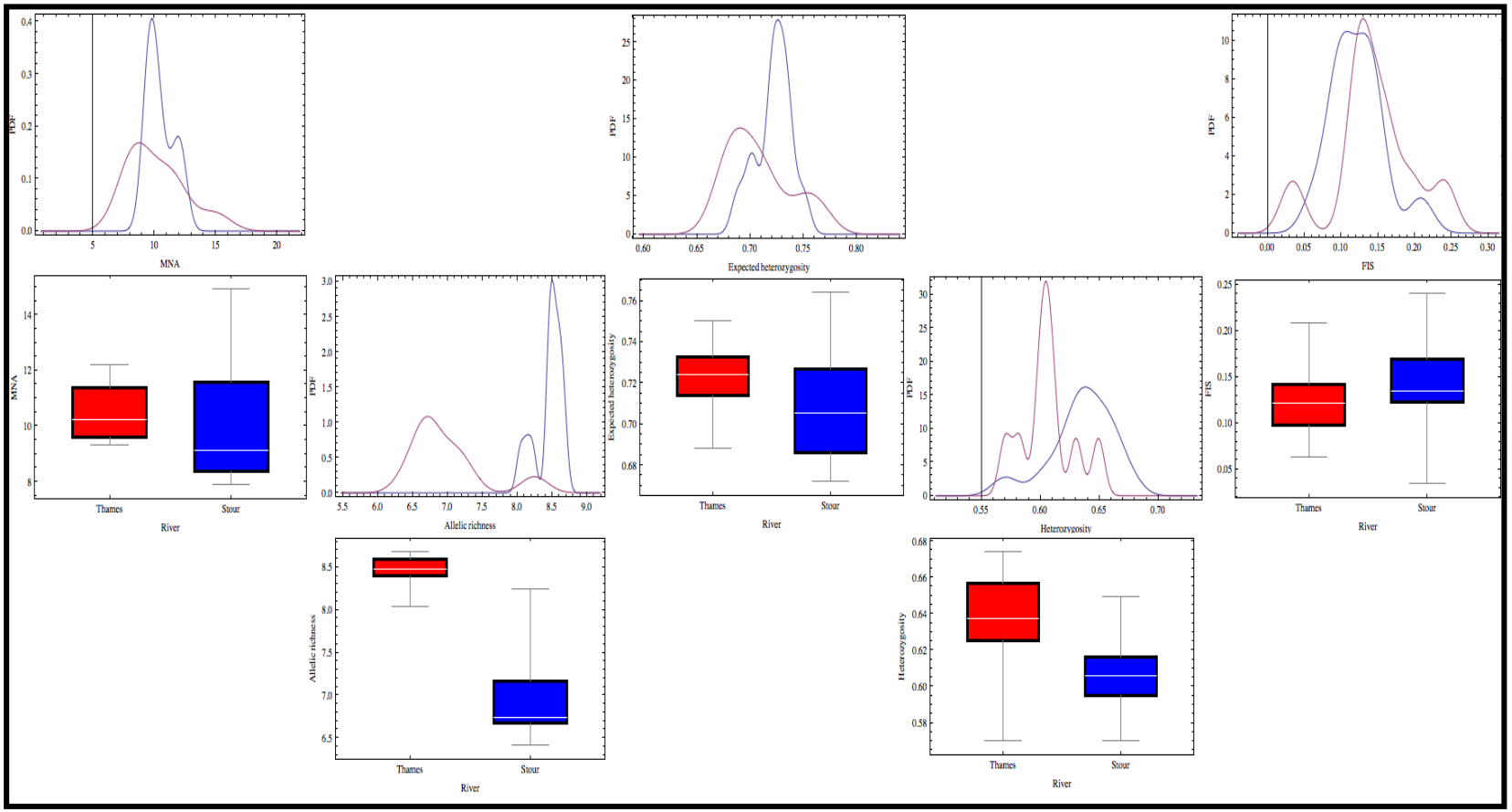


Figure 4.8: Comparison of mean levels of genetic diversity across sample sites between the Thames (red) and the Stour (blue) for: (left to right) mean number of alleles (MNA), allelic richness (A_r), gene diversity (expected heterozygosity), observed heterozygosity and F_{IS}. Top: Probability density function plots; Bottom: Box plots showing 50% quartile range, full range (bars) and median values (white line within box).

Table 4.10: Single-point estimates of the effective population sizes of the Thames and Stour Rivers' roach metapopulations, based on an analysis of 10 microsatellite loci.

Method	River						Reference
	Thames			Stour			
	N_e	Confidence Interval		N_e	Confidence Interval		
		2.5%	97.5%		2.5%	97.5%	
LD	1631	1261.9	2273.7	368	335.5	407.5	Hill 1981
LD_{cor}	1158.3	662.9	3445.2	1094.5	548.5	8440.2	Waples 2006

Table 4.10 lists the estimated effective population sizes for both the Thames and Stour populations as calculated by the two LD methods. The heterozygosity excess method resulted in estimates of infinity for both the Stour and for the Thames. Estimates of N_e for the Stour were calculated to be 368 (95% CI: 335.5-407.5) and 1094.5 (95% CI: 548.5-8440.2) for the uncorrected and corrected LD method, respectively. The estimates of N_e for the Thames were 1631 (uncorrected - 95% CI: 1261.9-2273.7) and 1158.3 (corrected - 95% CI: 548.5-8440.2).

4.3.4.2 Models of Microsatellite Evolution and Deviation from Migration-Mutation-Drift Equilibrium

Hardy's allele-size randomization test (2003) was employed to determine whether the microsatellite data are consistent with the null hypothesis that the impact of stepwise mutational pressure on genetic diversity and population divergence is negligible in comparison to inter-population migration rate and/or the inverse of the time since demographic isolation. Table 4.11 shows the global F_{ST}, R_{ST} and mean permuted pR_{ST} estimates for each locus and for all loci combined. With the exception of Lc27, Lco4 and Rru3, R_{ST} estimates for each of the loci are lower than the mean permuted values, with no loci showing a significant increase in value than expected from randomization of allele size classes. The hypothesis that the SMM model contributes significantly to the divergence estimates at these loci cannot be substantiated, at least within the time frame since the river populations last shared a common breeding population. As the value of R_{ST} is not greater than F_{ST}, and that a stepwise mutational process has not contributed to the discrepancy, one can assume that the R_{ST} measure does not indicate any significant biological difference. However, the test does assume a straight SMM model of microsatellite evolution and that populations are in migration-drift equilibrium.

Table 4.11: Results of Hardy's allele size randomization test.

Locus	F_{ST}	R_{ST}	pR_{ST}*	95% CI		*p*-value**
				2.5%	97.5%	
Lid1	0.0073	0.0056	0.0078±0.011	-0.0013	0.0390	0.5892
Lco4	0.0512	0.0695	0.0518±0.052	-0.0011	0.1734	0.6857
Rru3	0.0371	0.0080	0.0309±0.024	-0.0013	0.0781	0.2522
Ca1	0.0094	-0.0002	0.0099±0.014	-0.0013	0.0473	0.2308
Ca3	0.0028	-0.0013	0.0027±0.005	-0.0014	0.0176	0.1278
Ca12	0.0125	0.0048	0.0117±0.014	-0.0011	0.0499	0.4441
CypG3	0.0271	0.0038	0.0234±0.020	-0.0012	0.0682	0.1940
CypG48	0.0068	0.0050	0.0067±0.010	-0.0014	0.0345	0.5964
Lc27	0.0041	0.0117	0.0037±0.005	-0.0013	0.0136	0.9510
Lc290	0.0388	0.0035	0.0354±0.035	-0.0011	0.1197	0.2190
All	0.0198	0.0036	0.0140±0.009	0.0023	0.0348	0.0651

* Mean permuted R_{ST} values after 20000 allele size randomization permutations.
** 1-tailed tests, mean permuted R_{ST} < observed R_{ST},

The BOTTLENECK analysis indicated that in the Stour, eight out of ten loci exhibit an excess of heterozygotes relative to the expectation of heterozygosity under mutation-drift equilibrium under a model of strict IAM, whilst the converse was true under the TPM model. Assessing over all loci, the expected number of loci with heterozygote excess under IAM is 6.03, whilst the expected number of loci showing excess under the TPM model was 5.89. The Stour showed no evidence for systematic deviation from neutral expectations under the IAM (2-tailed p = 0.1739), however under the TPM there is a significant increase in loci showing heterozygosity deficiencies than expected under neutrality (2-tailed p = 0.0149). Similarly, in the Thames, there were a higher number of individual loci showing heterozygosity excess (7 out of ten) than expected under neutrality for the IAM model, although not significantly so (5.99, 2-tailed p = 0.3794), whereas the exact converse was true under the TPM (7 out of 10 loci showed heterozygosity deficiencies, against the neutral expectation of 5.85, although again not-significantly so (2-tailed p = 0.067)). The analyses do not support a significant reduction in population size, at least within 0.2 – $4N_e$ generations, under either the IAM or the TPM model of microsatellite evolution, within either the Thames of the Stour. The Stour, however, shows a significant deficit of heterozygotes under the TPM, as is similarly the case in the Thames (borderline significant), but not under the IAM.

The analysis of migration-drift equilibrium in the software 2mod was unequivocal: the model of migration and drift was supported 100%, whilst the model of drift alone was completely disregarded. To test whether contemporary levels of migration indicate the presence of true migration-drift equilibrium, the detection of migrants was carried out to assess for first generation and recent levels of effective migration (in GENECLASS (Piry et al. 2004) and BAYESASS+ version 1.01 (Wilson and Rannala 2003), respectively).

GENECLASS identified no first generation migrants whatsoever between the Thames and the Stour. Extremely low levels of inter-river migration between the Thames and the Stour were identified by BAYESASS+, such that migration rate into the Thames from the Stour was 0.00363±0.006, whereas the rate of migration from the Thames into the Stour was similarly low, 0.00552±0.006, although slightly higher. The results do not support a hypothesis of large-scale and ongoing migration between the two rivers.

4.3.4.3 Genetic Divergence Between the Thames and the Stour

Weir and Cockerham's θ was applied in the first instance to estimate levels of differentiation between the Thames and the Stour roach. After 1000 permutations, the value of 0.0198 was found to be statistically significant ($p < 0.001$). Additionally, standardized estimates of divergence were calculated: standardized F_{ST} (0.063), G'_{ST} (Nei's formulation, 0.02), G''_{ST} (0.074), D_{JOST} (0.056) and D_{NEI} (0.056). For both G'_{ST} and F_{ST}, and for D_{JOST} and D_{NEI}, their respective values are identical (0.2 and 0.056, respectively). Both F_{ST} and Nei's G_{ST}, non-standardized metrics, indicate the lowest degree of differentiation between the Stour and the Thames. The greatest amount of divergence is inferred from Meirmans & Hedrick's standardization of the latter statistic ($G''_{ST} = 0.074$).

4.3.5 Discussion

4.3.5.1 Levels of Genetic Diversity within the Thames and Stour

Table 4.12 lists indices of microsatellite diversity for five common and widespread cyprinids and one endangered cyprinid from the recent literature. The mean values of H_e in the UK roach are similar for those of similar small cyprinid species such as the minnow *Phoxinus phoxinus* (0.708) and the dace *L. leuciscus* (0.723), although values for H_o and F_{IS} are both lower in the roach, although this is probably due to pooling demes. Roach have greater levels of diversity than the tench, the endangered *C. lusitanicum* and the chub, with which it shares a similar level of F_{IS} in one study of within-drainage chub populations (roach = 0.213; chub = 0.203 (Dehais et al. 2010)).

Table 4.12: Levels of microsatellite diversity in a selection of widely distributed and endangered European cyprinids, including roach diversity data uncovered by this study.

Species	IUCN Listing	No. Loci	NA	H_o	H_e	F_{IS}	Reference
Rutilus rutilus	Least concern	10	20.50	0.618	0.740	0.213	Present study
Gobio gobio	Least concern	8	17.42	0.745	0.734	0.012	Blanchet et al. 2010
Phoxinus phoxinus	Least concern	8	20.38	0.706	0.708	0.021	Blanchet et al. 2010
Squalius cephalus	Least concern	10	8.85	0.595	0.588	0.019	Blanchet et al. 2010
Squalius cephalus	Least concern	5	-	-	0.656	0.203	Dehais et al. 2010
Leuciscus leuciscus	Least concern	15	13.57	0.721	0.723	0.022	Blanchet et al. 2010
Tinca tinca	Least concern	7	3.69	0.370	0.403	0.082	Kohlman et al. 2007
Chondrostoma lusitanicum	Critically endangered	6	7.17	0.291	0.280	-0.059	Sousa et al. 2008

NA = mean number of alleles; H_o = observed heterozygosity; H_e = expected heterozygosity; F_{IS} = population inbreeding.

Overall levels of roach genetic diversity are remarkably similar for both the Thames and the Stour 'populations', with directly predictable relationships between the two rivers in the number of alleles, number of effective alleles, observed heterozygosity, gene diversity and F_{IS}, although quite weakly so in the latter. The level of congruity in diversity indices is indicative of either retained ancestral diversity from a highly diverse common stock, or pronounced and ongoing gene flow. The single index in which the two populations differed unpredictably was in the number of private alleles. That the two populations differ in this respect suggests that the two rivers have diverged, and that this divergence is generally due to different mutational-drift trajectories, and that similarly high levels of present day diversity is a relic of prior connectivity in the recent geological past.

Both the Thames and the Stour populations exhibit a few locations where levels of heterozygosity deviate from null expectations. In almost all cases, locus-by-location comparisons exhibit lower heterozygosity relative to that expected from observed allele frequencies. All but four populations in the Thames show a significant reduction of heterozygotes relative to homozygotes over all loci (Molesey Weir Pool, Temple, Whitchurch and Buscot the exceptions), whereas all but two populations display similar excess of homozygotes in the Stour (Thurlow and Denham Mill being the exceptions) (see Tables 4.8 and 4.9 and Fig 4.7). Whilst the mean levels of F_{IS} and gene diversity were found to not significantly differ between the Thames and the Stour (although F_{IS} and H_e were higher in the Thames), mean observed levels of heterozygosity (H_o) were found to be significantly higher in the Thames than in the Stour (Fig 4.8). The frequency of locus-by-

location rejection of HWE (7.75% - 22.73%, Thames – Stour) is less than that found across 5 microsatellite loci in roach in one Swedish study (10/40 comparisons; 25% (Demandt & Bjorkland 2007)), but similar to that found in a later study of temporal differentiation in 5 microsatellite loci (2/26 comparisons; 7.6%, albeit before Bonferroni correction (Demandt 2010)).

A number of factors can impact upon levels of heterozygosity at neutral loci, most often expressed as the population-level inbreeding coefficient F_{IS}: most notably deviations from random-breeding (e.g. inbreeding (Conner and Hartl 2004)), the presence of age-structured populations and the sampling of distinct demes that have congregated in a single space. Sporadic and random deviations from HWE are expected with finite populations, and a low-level influence of null alleles cannot be entirely discounted, although their effects are probably minimal (Chakraborty et al. 1992; and see section 4.2.2.2). Such stochastic and technical artefacts probably explain the small number and non-patterned distribution of locus-by-location deviations from HWE in the Thames. These data are entirely consistent with known migration behaviour and individual dispersal of roach after the end of the spawning period, when roach migrate away from spawning areas to those richer in food to alleviate the loss of body weight by considerable reproductive effort and to prepare for next year's spawning run. Interestingly, the most downstream population at MWP is a site located immediately upstream of a heavily stocked segment of the River Thames, having thirty thousand fish roach released in the year 2000 (see below). It exhibits one of the smallest departures from HWE in the Thames (F_{IS} = 0.098, third lowest out of thirteen populations), which suggests that stocking of a small stretch of a major river may not necessarily result in impacting upon random breeding and local diversity if the diversity of the donor population is high (these fish were derived from private lakes (Nigel Hewlett, pers comm)). The introduced population may also be reproductively isolated if within MWP breeding preference is with other introduced roach, thereby negating localised Wahlund effects.

The frequency of locus-by-location deviations from HWE was almost three-fold greater in the Stour than in the Thames, despite a similar number of fish sampled overall. That locus Rru3, at five out of eight locations where it was genotyped, exhibits significant excess of homozygotes suggests that null alleles may be influential. A similar situation was observed in the European cyprinid tench (*Tinca Tinca*) whereby one locus, MTT-8, exhibited highly significant deviation from HWE in half of the studied wild populations, which the authors attributed to the influence of null alleles (Kohlmann et al. 2007). Loci

potentially affected by null alleles have also been identified in other freshwater fishes, e.g. burbot *Lota lota* (Elmer et al. 2008), perch *Perca fluviatilis* (Bergek and Björklund 2007), the bitterling *Rhodeus amarus* (Bryja et al. 2010) and the vairone *Telestes muticellus* (Marchetto et al. 2010). In most studies, and as observed in the present study, the inclusion of such loci seems to have little impact upon estimates of population structuring. Despite a small proportion of locus-by-location comparisons deviating from HWE, the overall effect of null alleles is not likely to be the main cause of deviation from HWE in the case of the Stour (only Rru3 showed any significant influence of null alleles with the EM test). As with the Thames, the Stour sampling occurred during post-spawning where shoals of roach tend to aggregate in communal feeding areas. Thus, it is probable that the Wahlund effect may explain heterozygote deficits within Stour samples. This is likely the case, for example, in the Anchor Bridge sample where 8/10 loci are significantly deficient for heterozygotes after application of correction. Anchor Bridge also shows a peak in both the mean number of alleles and allelic richness compared to other sites, consistent with a scenario in which more than one deme is represented synchronously.

Inbreeding is expected to increase over time within an isolated population of fixed size, and this process occurs at a greater rate in a smaller population than a larger one, even if mating is random within the populations. This may account for the significantly lower levels of observed heterozygosity recorded in the Stour in comparison to the much larger Thames drainage, albeit one that did not result in a significant difference in levels of F_{IS}. If population structure is more significant in the Stour (see section 4.4.4.2 below), individuals are more likely to spawn with a related individual in any given subpopulation. One final potential influence upon HWE is migration, in which migrating individuals possess alleles from donor populations whose frequencies do not match that of the recipient population. In highly vagile species, the presence of inter-population migration is expected to be high, and if population sub-structuring exists, the potential for violation of HWE is significant.

At mutation-drift equilibrium, the rate at which mutations are lost – or fixed – by genetic drift is offset by the introduction of new variants by mutation. In large populations, the time to equilibrium is longer than that in smaller populations, dependent on the mutation rate of the loci in question (Kimura 1983). Therefore the degree to which two populations differ in their allele frequencies is, at equilibrium, dependent on their effective population sizes.

4.3.5.2 Effective Population Sizes in the Thames and Stour

For both the Stour and for the Thames, the estimated size of the number of breeding individuals depended upon which method was used. The heterozygosity method proved entirely unsuitable for the roach, as an upper limit on the effective number of breeders could not be calculated. A similar situation has been observed in roach and perch (Demandt 2010) and other taxa (e.g. the natterjack toad *Bufo calamita* (Beebee 2009)).

The LD method is dependent upon sampling individuals from within contracted or fragmented populations, which in this case exhibit a non-random statistical association between non-linked allelic states (Hill 1981). However, longstanding inter-demic migration and / or past admixture events (including current admixture still present from daughter populations having once been part of a single interbreeding population) can also create persistent levels of LD, mimicking the effect of population contraction. The LD method assumes demographic isolation of a single, reproductively cohesive population unit. The LD method, then, may not be suitable for establishing the effective size of a single, large metapopulation. Even within a metapopulation structure, the potential occurrence of past or contemporary migration would not allow for an accurate representation of sub-deme effective population sizes. However, the bias correction of Waples (2006) allows for unequal sex ratios and variance in reproductive success. Using Hill's original formulation, the ratio of the effective size of the Stour to that of the Thames is 0.223, but with the correction applied the ratio decreases over four-fold to 0.995[37]. This decrease in the ratio may implicate a role of skewed reproductive success within the Stour that would hitherto have been overlooked. It has long been established in the roach that high concentrations of sewage effluents within streams can feminize juvenile and reproductive-age males (e.g. Jafri & Ensor 1976; Jobling et al. 2002; Nolan et al. 2001; and Bjerregaard et al. 2006). Recent experimental work on the determination of parentage success of both normal and intersex male roach indicate that highly feminized males are at a severe reproductive disadvantage, whilst 'mildly' feminized males are not (Harris et al. 2011). However, little has been done to uncover correlations between levels of genetic diversity and localized concentrations of feminizing agents, although the potential effect on population persistence has been modeled (An et al. 2009) (but see section 4.4).

[37] Assuming mutation-drift equilibrium, the IAM of Crow & Kimura (1970) relate gene diversity to theta by the equation $H = 4N_e\mu / 1+ 4N_e\mu$. Under these two assumptions, values of N_e for the Stour and Thames are 1185 and 1350, respectively, similar to those values given by Waples's correction (2006).

However, across both methods, the N_e of the Stour was found to be consistently less than that of the Thames. The simplest explanation is that this is the result of a lower census population size, but it could also come about as the result of a differential operating sex ratio (OST) at spawning, particularly if some males are compromised, within the Stour. Additionally, lower neighbourhood sizes (a function of dispersal ability and population density) within the Stour may reduce the effective population size. Population sub-division will also act to reduce overall estimates of N_e when sub-division is low (Pannell & Charlesworth 1999), but may increase it when the number of inter-connected demes is large and migration is limited (Robertson 1964) or when selective sweeps enable linked chromosomal variation to infiltrate other demes (Santiago & Caballero 2005). The preceding analysis considered – for initial simplicity –the Thames and Stour to be single, cohesive populations, when in all probability they represent metapopulations, with individual breeding units – connected by gene flow to a greater or lesser extent – located within their physical networks of rivers, streams and rivulets.

4.3.5.3 Adherence to Mutation-Migration-Drift Equilibrium

The analytical evidence of roach microsatellite variation, within each of the Thames and the Stour rivers, suggest that the levels of microsatellite diversity found within – and their divergence derivatives - are not biased by the strictly stepwise mutational process that early modelers of microsatellite evolution had proposed (see Valdés et al. 1993 and references therein). The most likely model of microsatellite evolution is pluralistic, invoking both strictly bidirectional stepwise mutations with a proportion of mutational events proportionate to the length of the microsatellite allele subject to mutation. The greater the length of the repeat motif, the greater the chance that slippage will occur and the greater the chance that slippage will encompass more than one repeat unit. Eventually, repeat length at each locus will reach equilibrium between length-associated mutations and point mutations breaking up arrays into loci of smaller length (Ellegren 2004). Pluralistic models of microsatellite mutation may more accurately represent microsatellite evolution than a strict SMM. Yue et al. (2006), in a study of microsatellite evolution in the carp *Cyprinus carpio* found, from tracking the inheritance of 49 distinct microsatellite loci, that the allele size class of a mutation differed from its parental progenitor by between -5 and + 2 repeat units. Similar occurrences of deviation from the strict SMM have been observed in salmonids (Angers & Bernatchez 1997). A reliance on determining allele coalescent events from the analysis of size classes in microsatellites based upon the strict SMM

model of evolution is flawed, as bias is introduced in the estimation of mutation rate and the relationship between IBS and IBD is oversimplified.

Even accounting for deviations from SMM, no signal of an overall reduction in population size was detected for either the Thames or the Stour populations, although both rivers showed a near-significant and significant (respectively) deficit in gene diversity under the TPM, consistent with an interpretation of a population expansion or the influence of the SMM upon microsatellite loci in populations with high current θ levels. High rates of gene flow and rapid population expansion can erode or falsely infer the presence of recent bottlenecking in species with low effective population sizes, respectively (Cristescu et al. 2010). A search of the literature does not reveal any instance in either of the two rivers studied here in which a whole metapopulation has suffered a significant decline or possesses exceedingly low allelic diversity relative to heterozygosity. Fish kills have occurred within both drainages as they have throughout the British Isles, due to a number of factors in addition to pollution, including disease and parasitism (e.g. Hewlett et al. 2009). Downstream and tidal regions of rivers are more prone to the cumulative effects of effluents and industrial and agricultural run-offs, with the Thames tidal zone having experienced high levels of industrial pollution with implications for fish species assemblages in the upper tidal areas (Araújo et al. 1999). That no obvious effect of bottlenecking was observed in the present study may be due to three reasons: i) there has been no decline in sufficient numbers as to manifest a signal in the gene pool; ii) there has been a serious decline in the past, but this occurred of the magnitude of $0.2\text{-}4N_e$ generations ago; iii) there has been a decline, but only locally, and when the data were pooled this signal was lost. Additionally, the EA have supplemented fish stocks in both rivers, which may have reinstated genetic diversity and so masked historical population declines (Peirson et al. 2001).

The programme 2Mod found 100% support for a model of migration and drift acting in concert, completely rejecting an isolation and drift only model of population genetic divergence. This result suggests one of two possible explanations: that not enough time has elapsed to establish equilibrium levels of divergence such that migration-drift equilibrium is assumed but is a relic of past connectivity; or secondly, that inter-population migration has occurred post-separation helping to mitigate against the erosive effects of genetic drift. However, 2mod assumes no, or a negligible, effect of mutation upon differentiation (i.e. mutational pressure has no effect on IBD in the drift model (assumes $\mu << 1/t$, where t = time in generations) and mutation rate is much smaller than migration

rate in the migration-drift model, such that all alleles inherited from migrants do not mutate in their new population), which, given the mutability of microsatellites, may not be an insignificant source of confounding influence. For instance, if one assumes an average mutation rate of roach microsatellite loci similar to that of the carp, of 5.56 x 10^{-4} per locus per generation (Yue et al. 2006), then μ is greater than 5.33 x 10^{-4}, or 1/1875 generations (elapsed since the Thames and Stour were last part of a single drainage). This will likely compromise the migration-drift model in which all new variation is assumed to derive from immigration. Assuming a stable population of 10000 individuals, one may expect 55.6 new mutations per generation. A new mutation that survives purging instantly adds to the degree of differentiation of one population from another.

The migration-drift model in 2mod places emphasis on the assumption that the rate of mutation is not greater than the rate of migration. The mean recent migration rate between the Thames and Stour as calculated by BAYESASS+ is 0.00457±0.006. This is around one order of magnitude greater than the average mutation rate of the carp, and assumed to be similar for the roach. It is unlikely that all alleles derived from migrants have remained unchanged through the linear process of inheritance, if migration has been constant over time. The relationship between F_{ST} and migration, assuming a Fisher-Wright model of equal population subdivision with equivalent effective population sizes and symmetric rates of migration (m), is directly linear, given by the equation:

$$F_{ST} = \frac{1}{4N_e m + 1}$$

If the BAYESASS+ rate of migration is a realistic representation of long-term rates of migration, long-term N_e for each river population is calculated to be approximately 2750 individuals. However, this ignores mutation, whose effect is generally considered negligible. Accounting for mutation, Wright's equilibrium equation (1951) becomes:

$$F_{ST} = \frac{1}{4N_e m + 4N_e m + 1}$$

Under this equilibrium model, an F_{ST} of 0.0198 is achieved with equal effective population sizes of approximately 2500 individuals. The failure to incorporate mutation into models of population divergence that have been physically separated for a significant number of generations may overestimate the degree of genetic similarity and overestimate their

respective effective population sizes. The mutability of microsatellite DNA should not be considered insignificant when other equilibrium processes are assumed.

It would be unwise to assume, based on little to no contemporary migration, the nature of microsatellite mutation rates and direction, and knowledge of the geological time that the populations became isolated, that the populations are at migration-mutation-drift equilibrium. The likelihood of populations having attained migration-drift equilibrium is lessened in temperate areas of the globe in which organisms have recently recolonized northern latitudes since the LGM, but is dependent upon N_e and the migration rate (Crow & Aoki 1984; Waples 1998). The results given by 2mod may be explained by reference to a violation of its assumptions (re: negligible rates of mutation), the assumptions of the Fisher-Wright model and/or not enough time has elapsed to achieve equilibrium. A naïve interpretation of an F_{ST} measure of differentiation is likely to underestimate the actual extent of genetic isolation exhibited by large and historically inter-connected populations.

4.3.5.4 Population Differentiation Between the Thames and Stour

All measures of genetic differentiation employed were statistically significant, although they differed on the extent of divergence indicated. Jost's D, which gave a figure of 0.056, is identical to Nei's estimation of standard genetic distance between the two rivers. As D is independent of population size, it essentially describes the influence of 'new variation' upon estimates of genetic divergence, independent of the action of genetic drift. Therefore, if one assumes that the two rivers have been demographically isolated during the entirety of the last 7500 years, then D describes the degree to which mutation has caused the two rivers' populations to differentiate across the microsatellite loci. The highest value of differentiation was given by Meirman & Hedrick's new estimator of F_{ST}, G''_{ST} (0.074). G_{ST} is best suited for small numbers of comparisons, perhaps explaining why the maximized estimate of θ resulted in a slightly lower estimate of differentiation (0.063).

Most estimators of population divergence generally make simplistic assumptions about the evolutionary properties of populations. The island-model of genetic structure is only an explanatory model, assuming infinite populations sizes, symmetric and constant rates of migration and an infinite number of subpopulations (Wakeley 2004). In fact, this idealized model may dramatically overestimate the degree to which populations are divergent, as the number of subpopulations is not generally accounted for, a criticism of other estimators of

F_{ST} based upon restricted sampling (e.g. G_{ST} (Rottenstreich et al. 2007)). Whitlock & McCauley (1999) have also cogently argued against strong inference based on this and other equilibrium models. The very suitability of F_{ST} needs to be reassessed if the effective sizes of roach exceed 1000 individuals per population, particularly if inferred from highly polymorphic loci (Neigel 1997) and gene flow is weak[38]. Although F_{ST} may still be useful as a simple descriptive metric for current purposes (baseline and minimum estimated allelic differentiation) and comparisons with prior published studies, its limitations should be made clear in any exposition of population structure, and an inference of migration rate or its calculation for populations with high effective population sizes from Wright's equations, or from traditional estimators, should be highly qualified or avoided where possible (Whitlock & McCauley 1999; Neigel 2002; Meirmans & Hedrick 2011).

4.3.5.5. Conclusions

The extant roach populations of the Thames and Stour were once part of a colonizing wave of roach with which they share a recent evolutionary history. Physical demarcation has existed for some 1875 roach generations. By adhering to a strict Wrightian model of population differentiation, upon which the axiomatic statistic F_{ST} is derived, one may conflate past connectivity with ongoing inter-population migration. Therefore, any statistic upon which migration-mutation-drift equilibrium is assumed is fraught with potential error. The roach populations of the two rivers are unlikely to have attained equilibrium given their high effective sizes and recent history of connectivity. The level of F_{ST} between the Stour and Thames is considered to be at the lower end of the scale of differentiation (0.0198) among many freshwater fishes, consistent with significant gene flow between them, either contemporaneously or in the past. A comparison of F_{ST} estimators is suggestive that the small amount of divergence observed (significant, < 8% at most) is due in the most part to the accumulation of new mutation and less so to drift, consonant with the levels of private alleles in physically isolated, large populations and the relative down-regulated efficiency of drift to remove variation. Past connectivity and / or current gene flow (both likely anthropogenic in nature) is likely to have contributed in some part to the low degree of differentiation observed, but this cannot be quantified. It

[38] Where N_e is expected to be large, the influence of the mutation rate on estimates of genetic divergence increase, particularly for F_{ST} (Neigel 2002). It is only when effective sizes are small (< 1000) that Wright's inclusive equation should be abbreviated to its familiar formulation, and mutation rate considered unimportant. This is consistent with the finding that a significant mutational effect is inferred for an estimated N_e of 2000 roach and above, but very little or no effect for populations of 1000 individuals or less.

must also be stressed that appropriate measures of differentiation should be used in concordance with prior knowledge of the species under investigation when conducting research on populations using genetic data from which so many informative parameters are extracted.

Regardless of what measure of differentiation was used, each estimate was statistically significant and therefore panmixia was completely rejected. The two rivers deviate enough from a random assortment of allelic diversity to warrant consideration as independent management units, even before further consideration of within-river differentiation. The next task is to ask whether there is significant deviation from panmixia within both rivers and assess the levels to which within-river locales differentiate with respect to their allele frequencies.

4.4 The Elucidation of Fine-Scale Genetic Structuring and Genetic Diversity within the River Suffolk Stour and the River Thames

4.4.1 Introduction

4.4.1.1 Biotic and Abiotic Influences of River Systems on Population Biology

River systems are perfect exemplars of habitats in which dispersal is physically limited to either one of just two directions (Fagan 2002), at least from the point of view of analytical tractability. The geometry of a river system not only affects the rate of directional migration between populations, but may also interact with life-history characteristics to influence the growth rate of that population (Goldberg et al. 2010). As briefly noted previously, passive dispersal may be severely biased due to the strong influence of flow directionality upon egg, larval and juvenile movements in river systems (Speirs & Gurney 2001). However, the ecological dynamics of inter-deme connectivity and metapopulation structure may also be dependent upon local as well as regional processes (distance-linked dispersal being an example of the latter). Brown & Swan (2010) found that local niche factors were more influential upon headwater community structures than were distance-decay relationships (e.g. IBD and a positive relationship between environmental similarity and distance), indicating that migration has little influence in headwater dynamics, at least in invertebrate communities. The population genetic consequences of increasing upstream isolation and a concomitant susceptibility to population declines, as a result of gene loss and a decrease in the ability to resist population contraction through an intake of migrants from downstream, are well characterized. Headwater populations are likely to experience recurrent bottlenecking, that, in spite of some regenerative migration, are likely to retain a signal of reduced population genetic diversity, with an excess of heterozygosity relative to allelic richness. Generally speaking, such populations should exhibit signs of genetic disequilibrium, depending upon the rate of immigration and the frequency and severity of population size contraction and re-expansion.

Habitat fragmentation in rivers may have a profound impact upon population dynamics and species assemblages along a river's length, particularly when severe impediments to migration are constructed (e.g. dams, weirs and other anthropogenic barriers). Humans

have had an impact on the mechanics of river regulation in the UK since the first century A.D., when modifications were implemented for land drainage and transportation (Sheail 1988). The creation of dams is an obvious barrier to river connectivity (Larinier et al. 2001). The construction of stream-crossings (road bridges and culverts) may also pose problems for fish passage if inappropriately constructed (Warren & Pardew 1998). Natural barriers to fish movements, like waterfalls, are also important determinants in the ecological dynamics of a river's fish population (Torrente-Vilara et al. 2011). Kruse et al. (1997) found that river volumetric flow along inclines of greater than 10% were prohibitive to the upstream movements of some salmonid species. Flourmills, which relied upon the construction of artificial waterfalls, were commonplace in the Middle Ages in the UK, with as many as 5000 mills recorded by the year 1086 (Sheail 1988).

Weirs pose a significant obstacle to migration and may fragment previously contiguous populations depending on their type - navigation, flow adjusting, flow gauging - and frequency. Navigation weirs were constructed alongside navigation locks in a great many of the world's rivers during the last 300 years to make rivers more navigable to an increase in industrial traffic along their courses. Due to an increased risk of flood damage to homes & property (increased due to domicile construction on floodplains and increased rates of run-off in urban areas), a huge number of hydrometric gauging-weirs have been installed throughout Europe & the UK (over 800 alone in England & Wales (White et al. 2006)). The installation of weirs can profoundly affect the constitution of fish communities (Poulet 2007). The installation of weirs has had a negative impact upon fish movements (White et al. 2006). Under low to moderate flows, flow-gauging weirs have a serious detrimental effect on the upstream migration of lampreys (Russon et al. 2011). Lucas & Frear (1997) found that the presence of flow-gauging weirs altered the migratory behavior of adult barbel below the weir, potentially separating reproductive schools into those that could pass the weir – and so migrated upstream – with those that could not –which dispersed downstream. Winter & Densen's (2001) longitudinal study of fish migration along a transect of the River Vecht in the Netherlands, interspersed with six weirs in the years 1960 – 1984, found that only 10 out of 32 species were able to ascend the weirs in 5-30% of the years, with larger species faring better during the winter months at downstream sites as compared to smaller species in which passing any weir was equally unsuccessful.

Many rivers are heavily modified with respect to their banks and side-channels. When the homogenization of this lentic area is of sufficient severity, this results in the loss of all suitable feeding, spawning and communal vegetation, thereby decreasing the likelihood of

migration along such 'corridors.' Further anthropogenic activities that may impact on the demographic integrity of a river's population include the process of abstraction, whereby water is removed from a stream either for irrigation purposes or to be diverted to geographically distal drainages with smaller catchments via a connective network of pipes. Abstraction may result in less than normal levels of flow, which may be particularly acute in affecting water levels with a combination of unseasonal rainfall patterns, modified embankments and high numbers of weirs, locks and mills.

Genetic studies of freshwater fish species have often found instances of genetic degradation and population isolation occurring as a result of channel obstruction or habitat fragmentation. Habitat patch size and the extent of demographic isolation are correlated with levels of allozyme diversity within populations of *Cottus gobio* within three catchments of central Europe (Hänfling & Brandl 1998). Hanfling & Weetman (2006) found that populations of the same species within the River Rye catchment of northeastern England displayed low levels of variation within the river, with levels decreasing with upstream location, suggesting that classic source-sink dynamics describes the genetic processes on the upper areas or that they retain the signal of historical colonization.

4.4.1.2 Hydrography of the Thames

The Thames is Britain's second longest river, after the Severn, and has a long history of association with mankind. The river originates some 300 metres above sea level as the Isis in the foothills of the Cotswolds, an area rich in springs that further develop into full-flowing tributaries joining up with the main Thames in the Thames Valley **(**Marsh and Lees 2003). The main river flows for 215 miles (346km) before discharging into the southwesterly North Sea, falling some 183 metres in the process (91.5 metres in its first 9 miles and 30.4 metres more in the following 11 miles (Ackroyd 2009)). From its source to London Bridge, the average gradient is 0.32m km^{-1} over 263km (Hughes & Willis 2000). The catchment of the Thames is the largest in Britain, covering an area of approximately 13000km^2 (Evans et al. 2003**),** with 9950km^2 of catchment above the tidal limit at Teddington (Hughes & Willis 2000) (see Fig 4.9)**.** The catchment area of the Thames has been radically altered from how it would have looked in Bronze Age Britain. Heavy urbanization and the agricultural conversion of land adjacent to the main stem and tributaries have greatly modified rates of overland run-off. The average run-off calculated from a period encompassing 1961-1990 was 250mm yr^{-1} (accounting for differences in

rainfall and evaporation (Johnson et al. 2009)). Increased run-off is compensated by extraction for significant public, agricultural and industrial usage. However the main river has suffered less due to abstraction than its tributaries, which has seen dwindling volumes of water contributing to the main river body volume. The average flow at the inter-tidal juncture at Teddington is 5.205 x 10^6 litres per day (Ackroyd 2009). Upstream of Didcot, the surrounding catchment is mostly rural, whilst downstream the main Thames flows through significant urban areas of Oxford, Reading, Slough and eastwards through greater metropolitan London en route to the Thames estuary. The non-tidal Thames is joined by a large number of tributaries and streams, but the major arteries are the Churn, the Thame, the Coln, the Leach, the Evenlode, the Cherwell, the Kennet, the Ver, the Mole, the Wey, the Medway, the Lea and the Roding (Ackroyd 2009).

Environment Agency data suggest that 60% of all River Habitat Surveys (RHS) conducted along the Thames record scores consistent with habitats that are significantly or severely modified (Johnson et al. 2009). Conversely, only 7% of surveys record pristine sites. All of the River Thames south of the lock at St John's is navigable (some 191 miles of river (Ackroyd 2009)) (see Fig 4.10) and has been adapted for large transportation vessels, with the construction of some 44 working and serviceable locks and weirs. The stretch of river between Oxford to the upstream village of Burcot was made navigable after an Act of Parliament in 1624, in which orders to construct locks and weirs was included, resulting in the first constructed lock at Iffley in 1630**.** From St. Johns lock in the west, the Thames is fully navigable to open sea. Hughes & Willis (2000) characterized and partitioned the main Thames channel into five zones according to the extent of modification: Zone A: 36km from source to the confluence with the River Coln, characterized by a natural channel with pool riffle channels, no weirs and diverse littoral habitat; Zone B: 38km stretch of river from the Coln confluence to Farmoor Reservoir, characterized by dredging and some channel modification, but the course remains natural with some littoral plant life, albeit less diverse than Zone A; Zone C: 79km stretch from Farmoor Reservoir to the confluence with the River Kennet. Here the Thames is heavily modified, including the presence of 16 weirs and in-channel and side channel dredging. There can be significant water abstraction at Farmoor Reservoir and Didcot power station. Habitat diversity is only moderate; Zone D: 87km from Kennet to tidal Thames at Teddington. This stretch is perhaps the most severely modified stretch of the entire river. 20 weirs service transportation along the river. Dredging and channel straightening are such that habitat diversity is low; Zone E: 23km to London Bridge. Here the river becomes more accessible and is less modified, although heavily urbanized.

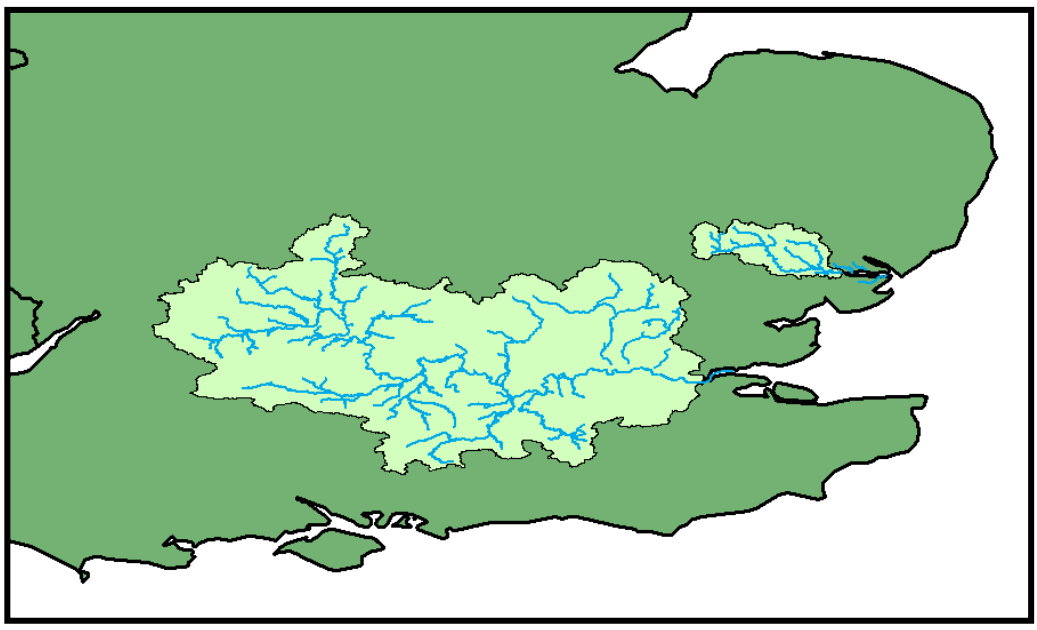


Figure 4.9: Map showing the difference in drainage area for the Thames and Stour (see text for details).

4.4.1.3 Hydrography of the Suffolk Stour

The Stour rises some 116m above sea level and flows for 76km forming much of the county boundary between Suffolk and Essex. The catchment area of the Stour, whilst not small for the Anglian region, only covers some 1036km^2 according to data from the UK Government (DEFRA 2010) (see Fig 4.8). The Stour catchment includes seven main tributaries in addition to the main trunk river. The surrounding land use is mostly agricultural: 83% of the Stour Catchment Flood Management area is agricultural (DEFRA 2010), with the river only flowing through one substantial urban town (Sudbury). The river falls some 58 metres over the first ten miles of its course, and by just 12 metres over the last ten miles. The underlying geology of the catchment is of chalk to the north and London Clay to the south, beneath a layer of semi-pervious boulder clay. 142 million litres of water are extracted daily at the water treatment facility at Langham, which extracts a number of agricultural run-off pollutants (e.g. metaldehyde, propyzamide, carbetamide and clopyralid (www.voluntaryinitiative.org.uk)). Water is also extracted from run-off and from groundwater, thus reducing the degree and frequency of heavy flow. The Stour is connected to the Great Ouse at Ely to the north near the source of the Stour (at Great Bradley) and the River Colne a few miles downstream at Wixoe via the Ely/Ouse transfer scheme, which, in addition to the Stour Augmentation Groundwater Scheme (SAGS), may affect water levels in the downstream portion of the river.

The Suffolk Stour was once widely used for industrial transportation of cargo, with some 24.4km of its length made navigable for large vessels (Fig 4.10). Recently, river enthusiasts have attempted to restore the Stour navigation downstream of Sudbury through to the mill at Bures. However, working locks are located downstream at Flatford and Denham, although they are not open during the winter when the sluices are opened for the anticipated increase in downstream water flow. Most of the 15 locks are now in a state of disrepair. Whilst there are few working locks, the Stour was employed as a source of energy for flour production through the construction of mills. One of the earliest records of a working mill occurred at the location that is now occupied by Bures Mill, in 1190. Between 1912 and 2009, the only navigable route upstream of Bures Mill has been either via the steep fall through the millstream itself, or through the sluice gates, which once posed a significant challenge even for the highly capable European eel (a fish pass has since been constructed). The river winds itself through tidal marshes after the Judas Gap - where a flood control defense barrier ("Fifty Six Gates") prevents the inundation of the Denham Vale with saltwater - to meet the estuarine surrounds of Cattawade before discharging into the North Sea at Harwich.

4.4.2 Hypotheses

Prior analyses have uncovered high levels of genetic diversity shared between both the Thames and the Stour, however the data suggest that the number of effective breeding individuals is lower in the Stour than in the Thames. The first investigative aim of this section is to determine why this difference exists. The most parsimonious explanation resides in the presumably lower census size of the Stour population than the Thames, resulting in a greater degree of relatedness among individuals. This situation will be exacerbated by the greater uniformity of suitable spawning habitat in the Stour than in the Thames, allowing for a greater dispersal of allelic diversity along the Stour. One may make the following prediction: there is likely to be some distance-component to the degree of genetic divergence among populations in the Stour, but less so – or absent – in the Thames, where long-distance dispersal may dominate.

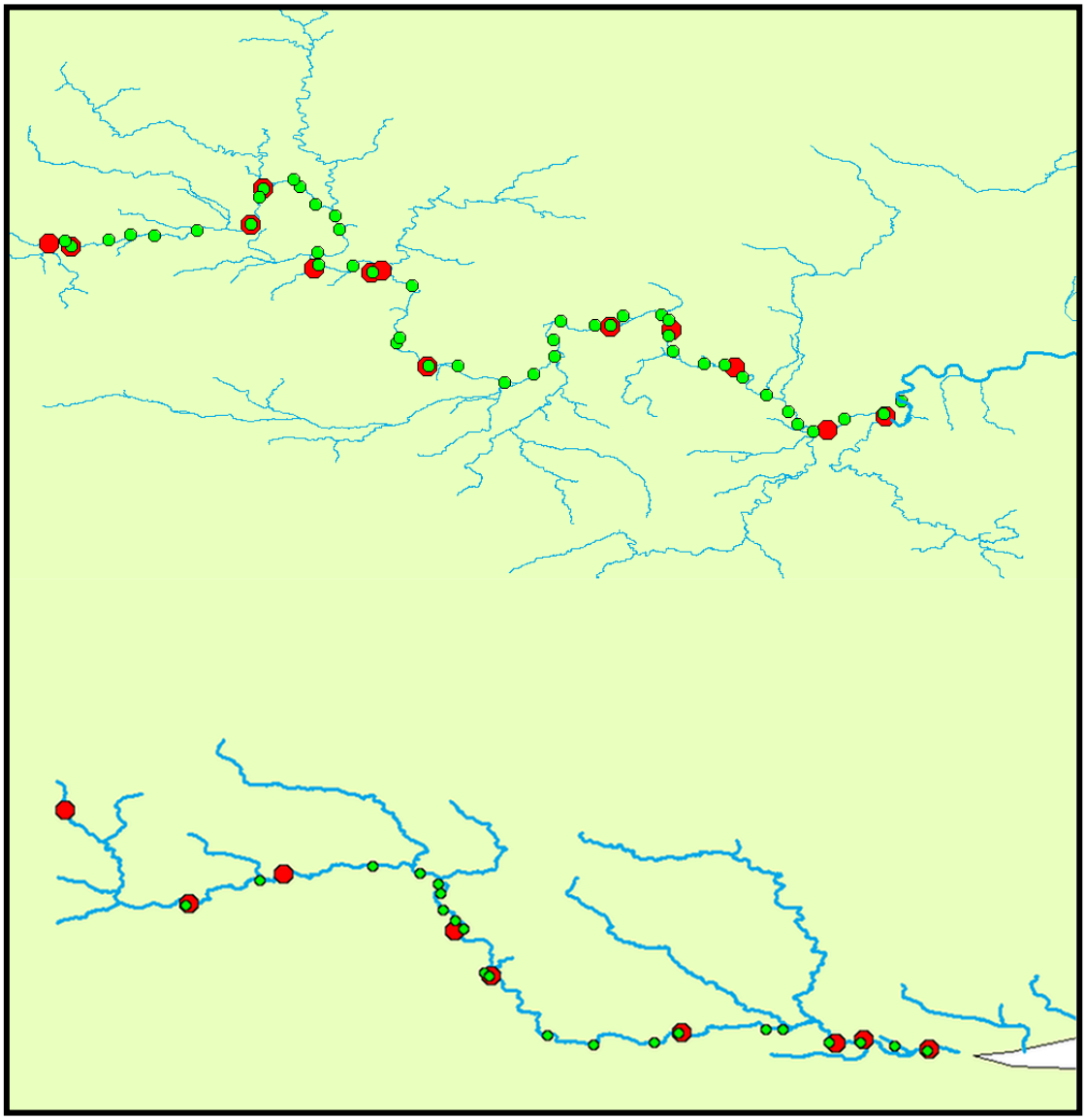


Figure 4.10: Potential barriers to gene flow. Map of anthropogenic constructions (locks, weirs and mills; green circles) found in the Thames (number of barriers = 45 in map; 44 between Ro and MWP) (top) and Stour (number of barriers = 21 in map and between Th and BL) (bottom). Red circles are sampled sites.

By sampling along a longitudinal transect, one can expand the above investigation of a genetic correlation between divergence and distance, by allowing a more robust test of putative environmental influences on the degree and extent to which populations differ genetically. Given the susceptibility of the Stour to low flows caused by low surface run-off and extraction for agricultural and domestic use, populations of fish in the Stour may be susceptible to the periodic and long-term influence of sewage effluents and their associated levels of endocrine disrupting chemicals (EDCs), particularly oestradiols and other feminizing chemicals. The physiological implications of EDCs in roach include downregulated milt production, reduced sperm density and motility compared to sites where EDCs are not at appreciably high levels (Jobling et al. 2002). The pernicious influence of feminizing hormones upon local genetic diversity and inter-population genetic

divergence may be readily tested. One would predict some relationship between genetic divergence and oestradiol concentrations in the Stour, but less so – or an absence of a relationship – in the Thames, with areas high in oestradiol concentrations possessing lower levels of roach nuclear diversity. Further, the influence of spawning habitat (lentic backwaters, tributaries) and barriers to migration may also influence the extent of inter-population genetic divergence with the expectation that correlation exists between genetic divergence between populations and the number of weirs, locks and tributaries between them.

4.4.3 Statistical Analysis

4.4.3.1 Within-Population Deviation from Equilibrium Conditions.

The program 2mod (Ciofi et al. 1999) was again utilized to determine whether individual populations had reached migration-drift equilibrium, and BOTTENECK (Piry et al. 1999) was used to uncover any signal of past demographic expansion.

4.4.3.2 Genetic Differentiation

As above, pairwise and mean estimates of divergence were calculated for all sampled sites within each of the two rivers.

In addition to the detection of genetic structuring among proscribed sampled locations, inferential detection of cryptic population structuring inherent in the data regardless of sampling location was also performed. The Bayesian software STRUCTURE (Pritchard et al. 2000) was used to uncover the number of 'true' populations, each of which conform to random mating and HWE, under a range of K populations (K1, K2…K*i*, where K*i* is the ith whole number of hypothetical populations), to see which hypothetical number is best supported by the data. Two approaches were used, both using the 'admixture' model of inference and correlated allele frequencies to account for recent evolutionary ancestry (Falush et al. 2003): firstly, no sampling information was used; secondly, because sampling information may be useful within populations with high levels of genetic diversity and low levels of inter-population divergence, sampling location information was inputted to inform the prior distributions of K (Hubisz et al. 2009). Up to 9 values of K

were inputted into the program ranging from 1 – 9 populations, and each K analysis was repeated 5 times (500000 MCMC iterations after 100000 burn-in) The most optimal K was calculated according to the methodology in Evanno et al. (2005), whereby the greatest rate of change in K (ΔK) between K populations is more informative as to the 'correct' K than by taking the K inferred from Pr (X|K) alone. The calculation of ΔK was carried out in the freely available online software STRUCTURE HARVESTER (Earl and von Holdt 2011). DISTRUCT version 1.1 (Rosenberg 2004), a program to display output from structuring programs graphically, was used to portray the most likely distribution of K populations within both the Thames and the Stour.

To determine the extent to which the variance in genetic diversity is apportioned hierarchically among individuals within sampled populations, among populations within rivers and between the Thames and Stour, an AMOVA analysis (Excoffier et al. 1992; Michalakis & Excoffier 1996) was conducted in the software Genodive. The F_{ST} analogue, ρ_{ST}, does not make assumptions about the underlying breeding system (Ronfort et al. 1998), and so was used here. Principal component analyses of multilocus genotypes can also be used to infer spatial patterning amongst sampled populations. Populations represent clusters of possible allelic and genotypic combinations found in a hyper-dimensional space, the dimensions of which are governed by the numbers of loci and the numbers of alleles found at each locus. A description of population similarity may be displayed graphically by plotting populations on axes that best explain the majority of the variance in differentiation among them. PCA was carried out in GenAlEx version 4.1 (Peakall & Smouse 2006), where the variance in the codominant genotypic distance (CGD, Smouse & Peakall 2006**)** was analyzed. Additionally, a PCA analysis was applied to distance estimates of allelic differentiation at all ten microsatellite loci for all 22 populations using pairwise estimates of unmodified F_{ST}.

4.4.3.3 Population Connectivity

In addition to genetic estimates of population structuring, of prime importance to wildlife managers is the determination of the origin and numbers of migrants. Genetic methods can only utilise neutral markers to determine population connectivity. Thus any loci found to deviate from neutrality via the above analyses were omitted from subsequent investigation of migration numbers and rates.

Inference of recent migration (within the last few generations) was utilised using the Bayesian software package BAPS 5.4 (Corander et al. 2004), using the admixture procedure described in Hänfling and Weetman (2006). Migrants are detected in BAPS by computing the posterior modal probability of the proportion of each individual's genotype arising from elsewhere in the sampled batch of populations. Following Hänfling and Weetman, an admixture model was applied and simulations were iterated for each individual 1000 times. By summing over all genomic proportions for each donor population and dividing by the recipient population size one was able to infer migration rate.

4.4.3.4 Landscape Genetics Analyses

To address if any environmental or physical factor has had an influence upon genetic divergence and inference of inter-population connectivity, suites of statistical tests were employed. The classic test is to determine whether equilibrium processes act along a geographically defined transect – in this case, along the course of the river. This isolation by distance effect is potentially greatest in one-dimensional habitats. The online software IDB (Jensen et al. 2005) was recruited to perform an initial analysis of isolation by distance for both the Thames and the Stour. The association of genetic distance/migration with distance was assayed by a Mantel test (Mantel 1967). All river distances were calculated by using the distance tool incorporated in the mapping software ArcGIS version 9.3 (ESRI Inc.) based upon hydrographic data collected by Moore et al. (1994).

Simple and partial mantel tests were then conducted between estimates of genetic divergence and a range of parameters, controlling for potentially confounding correlates in turn. These analyses were performed using the open source Mantel test resource ZT (Bonnet and Van de Peer 2002). The significance of Pearson's correlation coefficient, r, was assessed after 10^6 permutations of the data matrices. Further correlative analysis between measures of genetic diversity (e.g. effective population size, H_e, etc) and environmental factors (e.g. distance upstream from the tidal limit, mean level of oestradiol concentrations, number of weirs and tributaries, etc) were performed on the data, whilst controlling for the effect of sample size. Additionally, population means between the Thames and the Stour were screened for correct comparison by first determining differences in variances by Levene's test (1960). Depending upon the result of the Levene test, and upon whether the Stour samples exhibit a larger variance than the larger sample

of Thames populations (smaller sample sizes with larger variances are best tested non-parametrically (Zimmerman 1987)), either an independent t-test (alternatively, an unequal variance t-test where unequal variances are found in equally sampled sets of data (Ruxton 2006)) or a non-parametric Mann-Whitney U test was performed, all in SPSS version 17 (SPSS Inc).

IBD patterns exist due to the balance between genetic drift and the limited dispersal capacity of individuals. The scale of autocorrelation between individual genotypes may reflect the average extent of the dispersal capacity of individuals. The autocorrelation coefficient, *r*, (an analogue of Moran's *I* statistic) was calculated for all populations in both rivers as a function of predefined distance classes using PhiPT, a squared distance measure (Smouse and Peakall 1999). When *r* intercepts the x-axis (plotted on a spatial correlogram) there is no longer any statistically significant association between the genotypes of individuals sampled from populations that fall into that particular distance class. An autocorrelative approach was used, implemented in GenAlEx version 6 (Peakall and Smouse 2006), in which mulitallelic, co-dominant loci are assayed in a multivariate statistical system by combining alleles and loci so that the stochastic noise associated with correlative analyses is avoided (Smouse and Peakall 1999). Individual genetic distances were then correlated with geographic distances given by waterway distances between sampling sites. R was plotted as a function of seven distance classes in the Thames: 5km, 10km, 20km, 30km, 60km, 80km and 100km; and five classes in the Stour: 5km, 10km, 20km, 30km and 60km. Implementing Banks and Peakall's heterogeneity test, also in GenAlEx, one tested the significance of the resulting correlelograms.

One of the chief aims of landscape genetics is to associate genetic discontinuities with spatial data. The software BARRIERS version 2.2 (Manni et al. 2004) uses Monmonier's algorithm to correlate geographic information (XY co-ordinates) with distinct genetic disparities among populations. The method estimates Weir and Cockerham's θ and superposes this information onto a geographic map determined by XY coordinates whose 'edges' (potential 'barriers' between geographically close sites) are determined by Delaunay triangulation. The method was carried out for each locus separately (including any loci that behave non-neutrally) such that barriers supported by multiple loci would be identified. The top three supported barriers were calculated and displayed cartographically.

In a sister analysis to STRUCTURE, one again performed an analysis determining the presence of sub-populations, but this time incorporating actual spatial information in the form of geographical co-ordinates. This was done in the program Geneland (Guillot et al. 2005), an R-based Bayesian software package. Geneland apportions individuals to a pre-defined range of K subpopulations, whereas STRUCTURE allocates individuals sequentially for each K sub-population and determines the subpopulations with the highest likelihood. This latter approach may not identify the most optimal K (Excoffier and Heckel 2006). The Geneland MCMC simulations - of individual matrices consisting of genotypic and coordinate -data - were run 10^6 times with a thinning interval of 100. The inferred number of sub-populations was estimated from a range of 1-10 putative subpopulations. Allele frequencies were drawn from a Dirichlet distribution rather than from an F-model as the former model performs better than the latter (Guillot et al. 2005). Correlated allele frequencies were assumed (Guillot 2008). As recommended by Guillot and colleagues, the maximum value of the Poisson process was set to 100. Concordantly, the maximum number of nuclei used to construct the Poisson-Voronoi tessellation map was set at three times the maximum Poisson value, again as recommended by the software's authors.

4.4.4 Results

4.4.4.1 Population Equilibrium

The results of the 2mod analysis confirmed in all three of the independent runs that the populations within both the Thames and the Stour possess allele frequencies that most closely fit a model of population dynamics in which migration-drift equilibrium has been attained. Only on a few early iterations (post-burn-in) of the Markov Chain was the alternate model of drift-dominance favoured by the indicator variable *I*.

BOTTLENECK analyses indicated that none of the Thames samples were found to have significantly large excesses of heterozygosity relative to that expected under mutation-drift equilibrium, thus failing to reject the null hypothesis of demographically stable populations. However, significant heterozygote deficits were pinpointed at certain sites: MWP ($p = 0.0419$), Whitchurch ($p = 0.012$), Eynsham ($p = 0.009$) and Buscot ($p = 0.029$) (TPM model). In contrast to the Thames, four out of ten sites within the Stour displayed an excess of heterozygotes, suggesting some recent influence of population contraction: in the

tidal Brantham Lock (p = 0.0419), Mill Meadow (p = 0.0068), and the two most upstream sites of Stoke-by-Clare (p = 0.0161) and Thurlow (p = 0.0093). Only the TPM model suggested a significant excess of heterozygotes in the Stour, with any excess heterozygosities rejected at a statistical level under the model of the SMM.

4.4.4.2 Genetic Differentiation

Tables 4.13 and 4.14 show estimates of population differentiation between individual sampling sites within the Thames and the Stour. Generally, the G-test indicates that almost every population within both the study rivers are significantly different from all other populations (after Bonferroni correction, α = 0.05). In the Thames, the site at Eynsham loses significance with four sites downstream (Desborough Loop upstream to Temple) and Buscot upstream (Table 4.13). No other patterning is obvious. The remainder of unsupported divergence include Roundhouse with Northmoor, Buscot with Windsor, Northmoor with Whitchurch, and Desborough Loop with Whitchurch.

Table 4.13: estimates of genetic differentiation between pairs of sites within the Thames. Above diagonal: G-statistic significance (HS = highly significant (p << 0.001), shaded grey indicates significance lost after Bonferroni correction (p < 0.05)). Below diagonal: F_{ST} estimates (highlighted, underlined yellow values indicates significance after Bonferroni correction, whereas shaded grey indicates significance lost after Bonferroni correction).

	MWP	DL	Wi	Cl	Te	Wh	Do	Da	Cu	Ey	No	Bu	Ro
MWP		HS	HS	HS	HS	HS	HS	HS	HS	HS	HS	HS	HS
DL	0.062		<0.001	0.009	HS	HS	HS	HS	HS	0.156	HS	HS	HS
Wi	0.062	0.011		HS	HS	HS	HS	HS	HS	0.497	HS	0.005	HS
Cl	0.071	0.013	0.005		HS	HS	HS	HS	HS	0.017	HS	HS	HS
Te	0.070	0.021	0.006	0.013		HS	HS	HS	HS	0.003	HS	<0.001	HS
Wh	0.103	0.045	0.029	0.037	0.024		0.001	HS	HS	HS	0.077	HS	0.004
Do	0.097	0.036	0.028	0.037	0.018	0.005		HS	HS	HS	<0.001	HS	HS
Da	0.061	0.023	0.031	0.038	0.033	0.057	0.044		HS	HS	HS	HS	HS
Cu	0.109	0.053	0.041	0.044	0.021	0.014	0.011	0.066		HS	HS	HS	HS
Ey	0.060	0.004	0.001	0.005	0.005	0.040	0.028	0.024	0.039		HS	0.252	HS
No	0.092	0.041	0.025	0.027	0.019	0.006	0.01	0.056	0.019	0.025		HS	0.307
Bu	0.084	0.009	0.004	0.011	0.012	0.045	0.031	0.036	0.038	0.001	0.036		HS
Ro	0.098	0.046	0.037	0.040	0.028	0.009	0.005	0.048	0.015	0.033	-0.003	0.044	

In the Stour, all but one pairwise comparison is highly significant: Brantham Lock and Shalford Weir, although this is significant before conservative correction is applied (Table 4.14).

Table 4.14: estimates of genetic differentiation between pairs of sites within the Stour. Above diagonal: G-Statistic significance (HS = highly significant (p << 0.001), shaded grey indicates significance lost after Bonferroni correction (p < 0.05)). Below diagonal: F_{ST} estimates (highlighted, underlined yellow values indicates significance after Bonferroni correction, whereas shaded grey indicates significance lost after Bonferroni correction).

	BL	DM	St	AB	Sh	MM	RC	SbC	Th
BL		HS	HS	HS	0.02	HS	HS	HS	HS
DM	0.024		HS	HS	HS	HS	HS	HS	HS
St	0.007	0.014		HS	HS	HS	HS	HS	HS
AB	0.024	0.022	0.013		HS	HS	HS	HS	HS
Sh	0	0.026	0.011	0.020		HS	HS	HS	HS
MM	0.009	0.026	0.011	0.018	0.01		HS	HS	HS
RC	0.056	0.092	0.038	0.035	0.063	0.047		HS	HS
SbC	0.045	0.074	0.030	0.038	0.056	0.037	0.018		HS
Th	0.085	0.086	0.084	0.060	0.097	0.080	0.101	0.086	

Genetic differentiation, as measured by the fixation-index F_{ST}, was significant across all sites globally within both the Thames and the Stour (F_{ST} = 0.032 (95% CI: 0.008 – 0.060) and 0.039 (95% CI: 0.016 – 0.043), respectively; p < 0.05, 1000 permutations). Individual pairwise values of differentiation are displayed for each river in Tables 4.13 and 4.14. In both rivers, as indicated by the modest level of global differentiation, the level of genetic differentiation among sites is generally low, but significant. Of the 78 possible comparisons in the Thames, 59 are significant after Bonferroni correction. Of the 36 possible comparisons in the Stour, 32 are significant. However, some significant patterns were apparent. In the Thames, the site at Molesey Weir Pool (MWP) is significantly divergent from all other sites (mean F_{ST} = 0.081). MWP and Culham exhibit the greatest divergence (F_{ST} = 0.109). Although there is a general increase in divergence with upstream site, sites such as Days and Eynsham contravene the general trend by having scores (0.061 and 0.060, respectively) lower than sites Desborough Loop and Windsor just upstream of MWP. Sampling locations downstream of Culham generally show little divergence with their neighbouring sites. For example, Desborough Loop and the immediately upstream Windsor and Clivedon Island sites are minimally differentiated, as are Windsor and Clivedon Island and Temple; and Whitchurch and Dorchester. Eynsham and Buscot are less differentiated with Desborough Loop, Windsor, Clivedon Island and Temple, than either is with its immediate neighbour. The Stour has a number of sites whose differentiation is universally supported: Thurlow, Stoke-by-Clare and Rat's Castle are all well supported, as are Denham Mill and Anchor Bridge. Denham Mill aside, these sites are all located upstream of the others, and in the case of Thurlow, near the source of the river. Thurlow is also the most divergent population (mean F_{ST} = 0.085).

Mean estimates of total heterozygosity-corrected and standardized divergences over all populations were calculated in Genodive. Thames estimates: Mean G'_{ST} (Nei) = 0.032 (95% CI: 0.008 – 0.062); G''_{ST} = 0.125 (95% CI: 0.029 – 0.249); Jost's D = 0.093 (95% CI: 0.021 – 0.200). Stour estimates: Mean G'_{ST} (Nei) = 0.042 (95% CI: 0.022 – 0.064); G''_{ST} = 0.141 (95% CI: 0.073 – 0.223); Jost's D = 0.104 (95% CI: 0.048 – 0.179). All measures indicated strong statistical significance ($p << 0.001$, after 999 permutations). Pairwise estimates for the above metrics and standardized F_{ST} (θ) are shown in Tables 4.15 and 4.16 (Thames) and Tables 4.17 and 4.18 (Stour).

Table 4.15: Pairwise estimates of genetic differentiation within the Thames. Below diagonal: Standardized F_{ST} (F'_{ST}); above diagonal: Jost's D.

	MWP	DL	Wi	Cl	Te	Wh	Do	Days	Cu	Ey	No	Bu	Ro
MWP		0.169	0.167	0.196	0.186	0.294	0.284	0.158	0.305	0.145	0.252	0.214	0.275
DL	0.196		0.030	0.035	0.056	0.130	0.107	0.061	0.151	0.009	0.114	0.025	0.132
Wi	0.189	0.020		0.014	0.015	0.081	0.080	0.084	0.111	0.002	0.068	0.010	0.104
Cl	0.228	0.034	0.003		0.035	0.106	0.111	0.104	0.123	0.013	0.075	0.029	0.114
Te	0.208	0.055	-0.001	0.030		0.066	0.049	0.087	0.056	0.011	0.050	0.030	0.075
Wh	0.335	0.141	0.077	0.114	0.053		0.014	0.164	0.037	0.106	0.017	0.120	0.024
Do	0.317	0.113	0.076	0.119	0.032	-0.025		0.129	0.030	0.076	0.030	0.085	0.015
Days	0.187	0.066	0.092	0.122	0.090	0.189	0.142		0.182	0.060	0.154	0.093	0.134
Cu	0.348	0.175	0.122	0.143	0.051	0.025	0.015	0.218		0.102	0.050	0.097	0.042
Ey	0.174	-0.002	-0.015	0.006	-0.004	0.121	0.079	0.071	0.121		0.064	-0.003	0.088
No	0.286	0.125	0.062	0.078	0.035	-0.012	0.003	0.176	0.043	0.068		0.092	-0.007
Bu	0.230	0.015	-0.002	0.020	0.009	0.115	0.068	0.098	0.101	-0.020	0.083		0.115
Ro	0.319	0.144	0.104	0.128	0.066	0.002	-0.013	0.157	0.034	0.102	-0.038	0.121	

Table 4.16: Pairwise estimates of genetic differentiation within the Thames. Below diagonal: Standardized G_{ST} (G'_{ST}) (Nei); above diagonal: G''_{ST}. Nan = not analysed due to insufficient data available for summary statistical calculation, caused by missing data at a single locus.

	MWP	DL	Wi	Cl	Te	Wh	Do	Days	Cu	Ey	No	Bu	Ro
MWP		0.221	0.219	0.254	0.244	0.366	0.353	0.209	0.381	0.194	0.321	nan	0.346
DL	0.063		0.041	0.047	0.076	0.168	0.139	0.082	0.197	0.013	0.151	nan	0.172
Wi	0.063	0.011		0.019	0.021	0.107	0.106	0.112	0.147	0.003	0.092	nan	0.137
Cl	0.072	0.013	0.005		0.047	0.138	0.144	0.138	0.162	0.018	0.100	nan	0.149
Te	0.071	0.021	0.006	0.013		0.088	0.065	0.117	0.076	0.016	0.068	nan	0.100
Wh	0.102	0.044	0.029	0.036	0.024		0.018	0.211	0.050	0.142	0.024	nan	0.032
Do	0.097	0.036	0.028	0.037	0.017	0.005		0.167	0.040	0.101	0.040	nan	0.020
Days	0.061	0.023	0.031	0.038	0.033	0.057	0.044		0.236	0.082	0.201	nan	0.175
Cu	0.109	0.053	0.040	0.044	0.021	0.013	0.011	0.066		0.137	0.067	nan	0.056
Ey	0.058	0.004	0.001	0.005	0.005	0.039	0.027	0.024	0.039		0.088	nan	0.118
No	0.092	0.041	0.025	0.027	0.019	0.006	0.010	0.056	0.018	0.025		nan	-0.010
Bu	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan		nan
Ro	0.098	0.046	0.037	0.040	0.027	0.008	0.005	0.048	0.015	0.033	-0.003	nan	

For both datasets, θ is shown to understate the inferred level of divergence for standardized measures of F_{ST} (F'_{ST} and G''_{ST}). G'_{ST} (Nei) and D_{JOST} estimates are in general agreement with those determined using θ, except in the Thames where D_{JOST} estimates are similar to those of the standardized estimators, rather than traditional estimators of F_{ST}. In both the Thames and the Stour, G''_{ST} differentiates among populations to a greater degree than F'_{ST}.

Table 4.17: Pairwise estimates of genetic differentiation within the Stour. Below diagonal: Standardized F_{ST} (F'_{ST}); above diagonal: Jost's D.

	BL	DM	St	AB	Sh	MM	RC	SbC	Th
BL		0.026	0.009	0.012	0.006	0.010	0.048	0.038	0.083
DM	0.056		0.015	0.014	0.039	0.031	0.097	0.074	0.090
St	-0.007	-0.016		0.009	0.013	0.019	0.043	0.033	0.084
AB	0.026	0.020	-0.015		0.020	0.012	0.050	0.037	0.071
Sh	-0.016	0.024	-0.032	0.009		0.020	0.050	0.043	0.080
MM	0.012	0.048	0.009	0.014	-0.001		0.043	0.030	0.080
RC	0.112	0.195	0.083	0.121	-0.002	0.052		0.001	0.104
SbC	0.088	0.149	0.031	0.085	0.031	0.038	-0.089		0.084
Th	0.229	0.229	0.097	0.201	0.099	0.190	0.276	0.204	

Table 4.18: Pairwise estimates of genetic differentiation within the Stour. Below diagonal: Standardized G_{ST} (G'_{ST}) (Nei); above diagonal: G''_{ST}. Nan = not analysed due to lack of data at one locus. Nan = not analysed due to insufficient data available for summary statistical calculation, caused by missing data at a single locus.

	BL	DM	St	AB	Sh	MM	RC	SbC	Th
BL		0.081	0.028	0.040	0.021	0.033	nan	0.124	0.264
DM	0.026		0.050	0.046	0.125	0.104	nan	0.243	0.283
St	0.009	0.015		0.033	0.045	0.067	nan	0.116	0.286
AB	0.012	0.014	0.009		0.068	0.042	nan	0.129	0.242
Sh	0.006	0.039	0.013	0.020		0.068	nan	0.145	0.256
MM	0.010	0.032	0.019	0.012	0.020		nan	0.103	0.267
RC	nan	nan	nan	nan	nan	nan		nan	nan
SbC	0.037	0.073	0.033	0.037	0.043	0.030	nan		0.283
Th	0.083	0.089	0.085	0.072	0.080	0.080	nan	0.084	

Fig 4.11 shows the relationship between the change in ΔK with an increase in hypothesized random-mating demes as determined from the allelic data for all loci and all individual roach considered with prior location information (right) and without (left) (lower graph). The upper graphs display the mean log probability of the data supporting each particular hypothetical number of populations (K). The STRUCTURE analysis showed that K = 2 when sampling location was not take into account, but K = 5 when sampling information was inputted. For K = 2, most individuals' genotypes in each river are composed of one of the two hypothesized contributing gene pools, which are roughly

co-distributed with the two rivers' individuals, with the exception of Thurlow (Stour) which has a closer affinity to the Thames (Fig 4.12 upper).

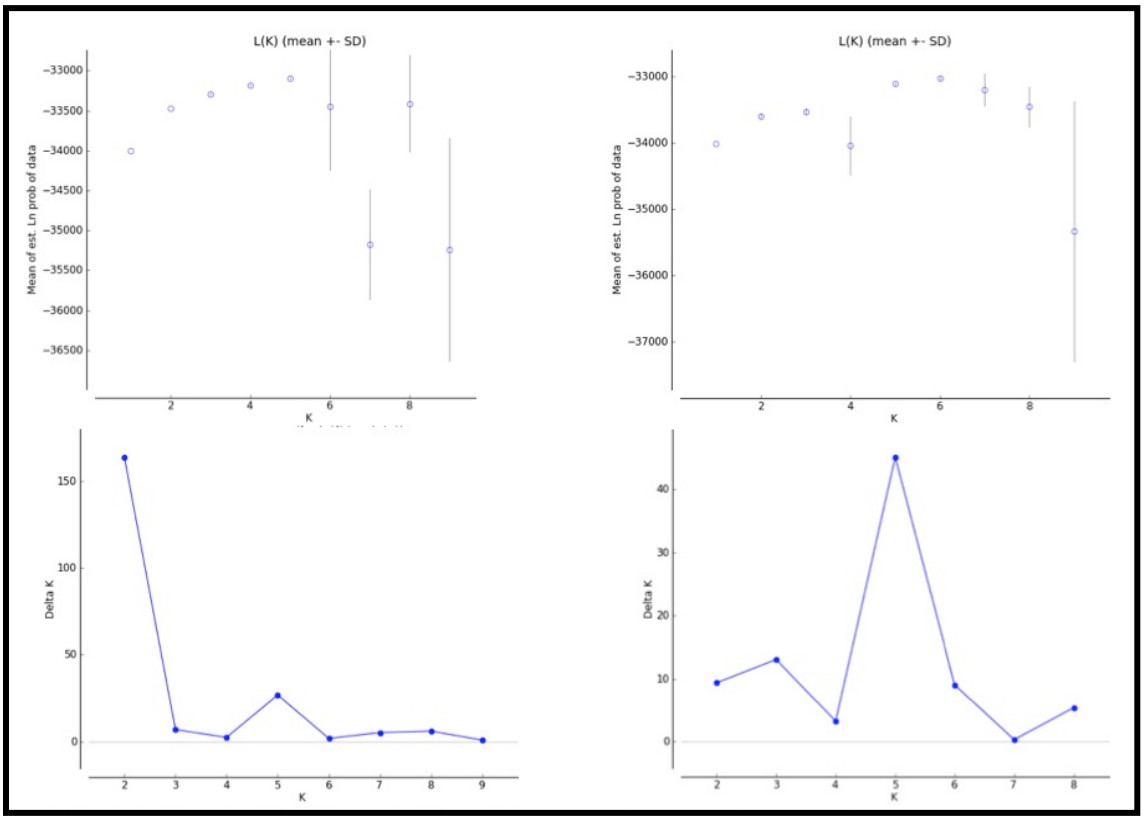


Figure 4.11: Cryptic population structure analyses. Optimal number of HWE populations present within the entire dataset as a whole, as deduced by assessing change in ΔK outputted by STRUCTURE analysis. The left hand analysis does not adopt sampling information, whereas the second hand analysis does. See text for details.

Fig 4.12 (lower) reveals the genomic proportions of each individual as can be apportioned to one of five HWE populations. The Thames (brown/blue) and Stour (green/yellow) populations as belonging to distinct rivers are more clearly defined. Both rivers display one possible remnant and shared population. The populations of Thurlow (Stour) and Molesey Weir Pool (Thames) are composed of individuals whose genomes exhibit an almost unique genetic signal.

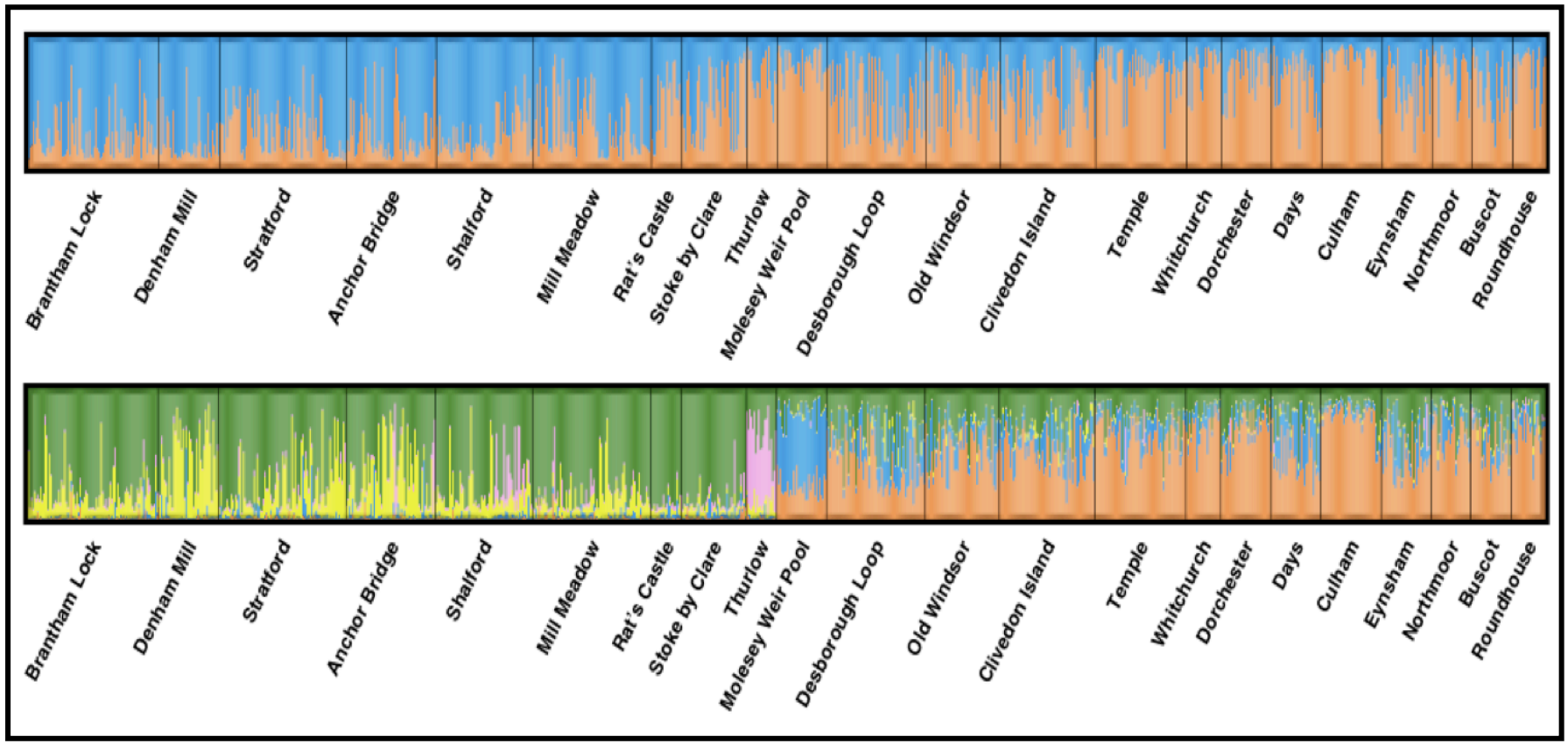


Figure 4.12: Pictorial representation of STRUCTURE results. Above: K populations = 2 when no geographic information is used. Below: K populations = 5 when prior information is utilized. Each diagram shows the proportion (Q) of each individual's genotype – represented as a vertical line - that is assigned to each of K hypothesized HWE populations. Brantham Lock – Thurlow = Stour; Molesey Weir Pool – Roundhouse = Thames. The colours utilized in the two figures do not correspond. Black vertical lines delineate batches of individuals belonging to each labeled population.

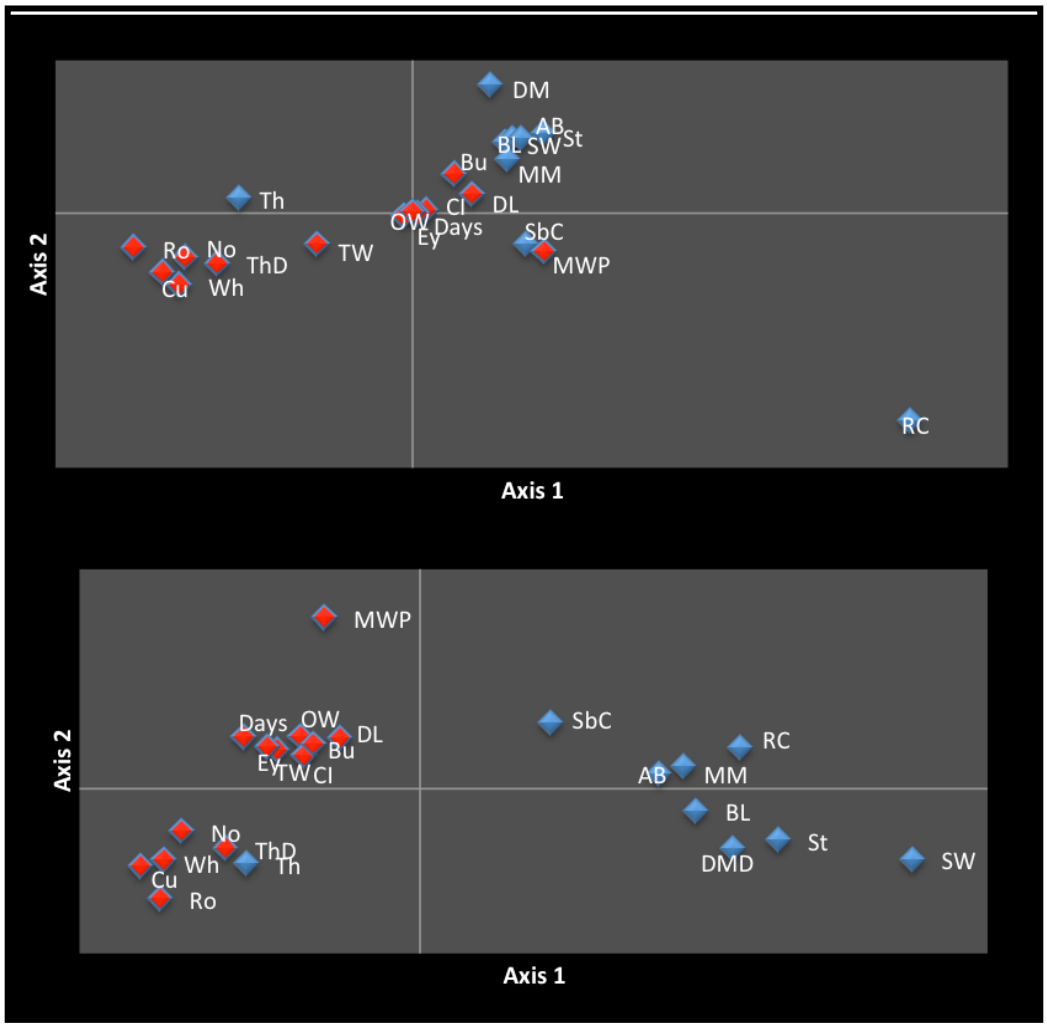


Figure 4.13: PCA graphs describing the variance in allele frequencies (F_{ST}, upper panel) and genotype frequencies (GCD, lower panel) that can be apportioned between the first two axes of variation for all 22 populations of roach. Red populations are found within the Thames and blue populations are found within the Stour.

Differentiation among all 22 populations from both the Stour and the Thames were visualized using PCA (Fig 4.13) on both allelic data (based on a F_{ST}-distance matrix) and upon the distribution of pairwise data on co-dominant genotypic distances (GCD). In both plots, the first two axes explain most of the variation inherent in the two distance-based methods (first two axes account for 50.6% of the variance in the F_{ST}-matrix, whereas the genotypic-distance matrix represents 65.57% of the cumulative variance). The genotypic distances better delineate the two river populations into coherent aggregates, primarily due to variation on the first axis, with the exception of Thurlow, which groups in amongst five upper and middle course Thames populations. Fig 4.13 (bottom) suggests that in the Thames, Molesey Weir Pool is isolated with respect to the other 12 populations. The genotypic data also suggest the existence of two putative clumps within the Thames: one

consisting of two of the three most upstream populations (Roundhouse – the most upstream population – and Northmoor) and three contiguous 'middle river' populations (Wh, ThD and Cu)); and another consisting of populations Desborough Loop to Clivedon Island from the lower-middle Thames, plus Days and the upper Thames population of Buscot. The GCD PCA analysis of the Stour populations separates the headstream population of Thurlow from all others. However, little patterning is evident elsewhere within the Stour, except that each population appears to reside separately in the space described by the two axes. The F_{ST} PCA results are similar to the GCD results in that MWP and Thurlow are shown to be genetically distinct, and that the two clumpings in the Thames are present, although less spatially obvious, except that the sample from Temple Weir (TW) is placed equidistantly between them. Furthermore, locations Stoke by Clare and Rat's Castle in particular – two contiguous upstream Stour populations – deviate from the non-Thurlow populations in the F_{ST} analysis, although this is less obvious from Fig 4.13 (bottom).

Table 4.19 unveils the results of assignment tests adopted in Geneclass 2.0. Panmixia was rejected for both rivers, as correct assignment was high in the Stour (89.1%) and greater than random in the Thames (61.9%), based on Rannala & Mountain's Bayesian model (1997). Overall, each individual roach was correctly assigned to its population of origin 75.3% of the time. However, the variation of correct assignment was much greater in the Thames (mean % correct assignment 64.0±16.43) than found in the Stour (mean % correct assignment 88.8± 5.11). The range of correct assignment ranged from 80% at Rat's Castle to 95% of Thurlow roach in the Stour, and 38.5% at DL to 91.3% at Whitchurch in the Thames).

The distribution of microsatellite diversity among individuals, populations and rivers was hierarchically quantified by applying an AMOVA analysis (Table 4.20). At all levels above that of the individual, the hierarchical distribution of genetic diversity was highly significant ($p < 0.01$), although most genetic diversity was found within and among individuals within populations (96.1 %, F_{IT} = 0.153, 95% CI: 0.113-0.196; F_{IS} = 0.114, 95% CI: 0.017-0.090). However, 3.8% of the variance in allele frequencies was to be found among sup-populations within each of the Thames and Stour (2.5%, F_{SC} = 0.026, 95% CI: 0.010-0.045) and between the two rivers themselves (1.3%, F_{CT} = 0.013, 95% CI: 0.006-0.022).

Table 4.19: Results of the assignment tests as implemented in Geneclass vs. 2.0

		Site	Reference Population	
			Stour	
			Correct	Incorrect
Assigned Sample	Stour	BL	87.2	12.8
		DM	92.5	7.5
		St	90.5	9.5
		AB	91.5	8.5
		Sh	92.2	7.8
		MM	88.5	11.5
		RC	80.0	20
		SbC	81.4	18.6
		Th	95.0	5.0
		Mean	88.8	11.2
		Overall	89.1	10.9
			Thames	
			Correct	Incorrect
	Thames	MWP	60.6	39.4
		DL	38.5	61.5
		OW	65.3	34.7
		Cl	52.4	47.6
		Te	76.7	23.3
		Wh	91.3	8.7
		Dor	72.7	27.3
		Days	51.5	48.5
		Cu	82.5	17.5
		Ey	42.4	57.6
		No	76.9	23.1
		Bu	48.1	51.9
		Ro	72.7	27.3
		Mean	64.0	36.0
		Overall	61.9	38.1
	Global	Mean	75.5	24.5
		Overall	75.3	24.7

Table 4.20: AMOVA results.

Source of Variation	%Var	*F*-stat	*F*-value	Std.Err.	95% CI		*P*-value
					2.5%	97.5%	
Within individual	0.847	F_{IT}	0.153	0.022	0.113	0.196	-
Among individuals within populations	0.114	F_{IS}	0.119	0.017	0.090	0.154	0.001
Among populations within rivers	0.025	F_{SC}	0.026	0.010	0.010	0.045	0.001
Among rivers	0.013	F_{CT}	0.013	0.004	0.006	0.022	0.001

4.4.4.3 Population Connectivity

Tables 4.21 and 4.22 show the migration rate into (recipient) and out of (donor) each sampled population within the Thames and Stour, respectively. From Table 4.21, MWP seems to have received no recognizable migrants over the preceding few generations. It has, however, produced migrants particularly to its nearest neighbours, but also to the most upstream site at Roundhouse, although the mean rate of migrants deriving from the area is the lowest in the Thames, 0.00208±0.00259, marking it as the lowest contributing population in the sampled Thames. Fig 4.14 better shows the pattern of both immigration (top) and emigration (bottom) by plotting mean values against location. There is a pattern of a decline in immigration with linear position upstream within the Thames, although this is not statistically significant (Pearson's r = 0.352, 2-tailed p-value = 0.238). However, if MWP is removed statistical significance is attained (Pearson's r = 0.672, 2-tailed p-value = 0.012).

Conversely, whilst it is generally observed that levels of emigration from a population increase with upstream positioning, it is not statistically significant (Pearson's r = 0.293, 2-tailed p-value = 0.331). The most active contributors of migrants are the two populations found directly upstream of the confluence of the Thames (Days) and Thame (Dorchester) (mean values: 0.02241±0.00225 and 0.02461±0.02475), respectively. The two most downstream populations contribute the least amount of migrants (MWP: lowest contributor: mean rate of 0.00208±0.00259; DL: mean rate of 0.00282±0.00459). Overall, the rate of migration was 0.00993± 0.04919 averaged across all populations.

Overall migration rate in the Stour, however, was an order of magnitude greater (mean migration rate of all populations = 0.09650±0.02418) than the global average found in the Thames. The vast majority of migrants in the Stour seem to derive from the Shalford Weir area (Table 4.22 and Fig 4.15), with a mean value of 0.11330±0.04263. Like the Thames, there seems to be a decrease in the rate of immigration with location upstream, although this is statistically unsupported (Pearson's r = 0.362, 2-tailed p-value = 0.337). This value approaches greater significance if the two sites of Brantham Lock and Denham Mill – the former of which is found at the tidal limit of the Stour - are eliminated (Pearson's r = 0.636, 2-tailed p-value = 0.124).

Table 4.21: BAPS inference of recent migration rates (based on eight microsatellite loci) – as a proportion of recipient individual genotypes purported to derive from elsewhere, weighted by recipient population size - from putative donor populations to putative recipient populations within the Thames. Migration rates above 0.01 migrants per generation are underlined.

		Donor															
		MWP	**DL**	**OW**	**CI**	**Te**	**Wh**	**Dor**	**Days**	**Cu**	**Ey**	**No**	**Bu**	**Ro**	**Σ**	**Mean**	**St. Dev**
Recipient	**MWP**		0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
	DL	0.00785	-	0.01492	0.00954	0.01154	0.02031	0.05892	0.02431	0.01185	0.01615	0.00492	0.04015	0.00846	0.22892	0.01908	0.01574
	OW	0.00143	0.00122	-	0.00306	0.00367	0.02041	0.03020	0.00673	0.00061	0.00163	0.00020	0.02388	0.00878	0.10184	0.00849	0.01039
	CI	0.00429	0.00794	0.01841	-	0.00286	0.02524	0.05032	0.02349	0.01190	0.02206	0.03254	0.00746	0.01175	0.21825	0.01819	0.01369
	Te	0.00167	0.00133	0.01567	0.00033	-	0.00800	0.03200	0.03683	0.03117	0.01533	0.00233	0.00950	0.00383	0.15800	0.01317	0.01325
	Wh	0.00478	0.00391	0.04391	0.02217	0.00174	-	0.00000	0.02522	0.00043	0.00826	0.04217	0.00000	0.01391	0.16652	0.01388	0.01603
	Dor	0.00424	0.01545	0.05364	0.00121	0.00636	0.00061	-	0.08939	0.01848	0.03879	0.01152	0.00121	0.01697	0.25788	0.02149	0.02679
	Days	0.00000	0.00091	0.00000	0.00030	0.00394	0.00030	0.00121	-	0.00000	0.00000	0.00000	0.01970	0.00000	0.02636	0.00220	0.00563
	Cu	0.00000	0.00250	0.00350	0.00300	0.00450	0.00700	0.04525	0.04050	-	0.00825	0.00825	0.01375	0.00800	0.14450	0.01204	0.01487
	Ey	0.00030	0.00061	0.00091	0.00000	0.01727	0.01788	0.03788	0.02515	0.00030	-	0.00000	0.04000	0.00242	0.14273	0.01189	0.01536
	No	0.00000	0.00000	0.00000	0.00000	0.02077	0.00154	0.01269	0.00000	0.00000	0.00000	-	0.00154	0.00000	0.03654	0.00304	0.00665
	Bu	0.00000	0.00000	0.00593	0.00000	0.00000	0.00000	0.00000	0.02370	0.00000	0.00111	0.00296	-	0.00185	0.03556	0.00296	0.00678
	Ro	0.00045	0.00000	0.00045	0.00773	0.00000	0.01091	0.00045	0.00000	0.00045	0.00045	0.00000	0.01136	-	0.03227	0.00269	0.00449
	Σ	0.02501	0.03388	0.15734	0.04735	0.07265	0.11219	0.26893	0.29533	0.07521	0.11205	0.10490	0.16855	0.07598			
	Mean	0.00208	0.00282	0.01311	0.00395	0.00605	0.00935	0.02241	0.02461	0.00627	0.00934	0.00874	0.01405	0.00633			
	St. Dev	0.00259	0.00459	0.01809	0.00658	0.00690	0.00940	0.02248	0.02475	0.01011	0.01199	0.01401	0.01437	0.00584			

Table 4.22: BAPS inference of recent migration rates (based on eight microsatellite loci) – as a proportion of recipient individual genotypes purported to derive from elsewhere, weighted by recipient population size - from putative donor populations (columns) to putative recipient populations (rows) within the Stour. Migration rates above 0.01 migrants per generation are underlined.

		Donor											
		BL	**DM**	**St**	**AB**	**Sh**	**MM**	**RC**	**SbC**	**Th**	**Σ**	**Mean**	**St Dev**
Recipient	**BL**	-	0.00419	0.01140	0.00047	0.05872	0.00419	0.01140	0.00616	0.01012	0.10663	0.01333	0.01875
	DM	0.00000	-	0.00650	0.00000	0.10925	0.00025	0.00075	0.00225	0.00025	0.11925	0.01491	0.03818
	St	0.11095	0.07393	-	0.01655	0.12964	0.00762	0.01631	0.03964	0.01214	0.40679	0.05085	0.04813
	AB	0.02746	0.01102	0.03898	-	0.11763	0.00305	0.00288	0.02644	0.00390	0.23136	0.02892	0.03831
	Sh	0.00000	0.00000	0.00000	0.00000	-	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
	MM	0.03769	0.02949	0.02090	0.00423	0.09859	-	0.00744	0.00564	0.00564	0.20962	0.02620	0.03183
	RC	0.00000	0.00000	0.00000	0.00000	0.03350	0.00000	-	0.00400	0.00000	0.03750	0.00469	0.01173
	SbC	0.01186	0.00023	0.00767	0.00023	0.04256	0.00023	0.00186	-	0.00840	0.07305	0.00913	0.01423
	Th	0.00000	0.00000	0.00100	0.00000	0.08000	0.00150	0.00100	0.00000	-	0.08350	0.01044	0.02811
	Σ	0.18796	0.11885	0.08645	0.02148	0.66989	0.01684	0.04163	0.08414	0.04045			
	Mean	0.02350	0.01486	0.01081	0.00268	0.08374	0.00210	0.00520	0.01052	0.00506			
	St Dev	0.03819	0.02593	0.01339	0.00579	0.03583	0.00272	0.00595	0.01452	0.00482			

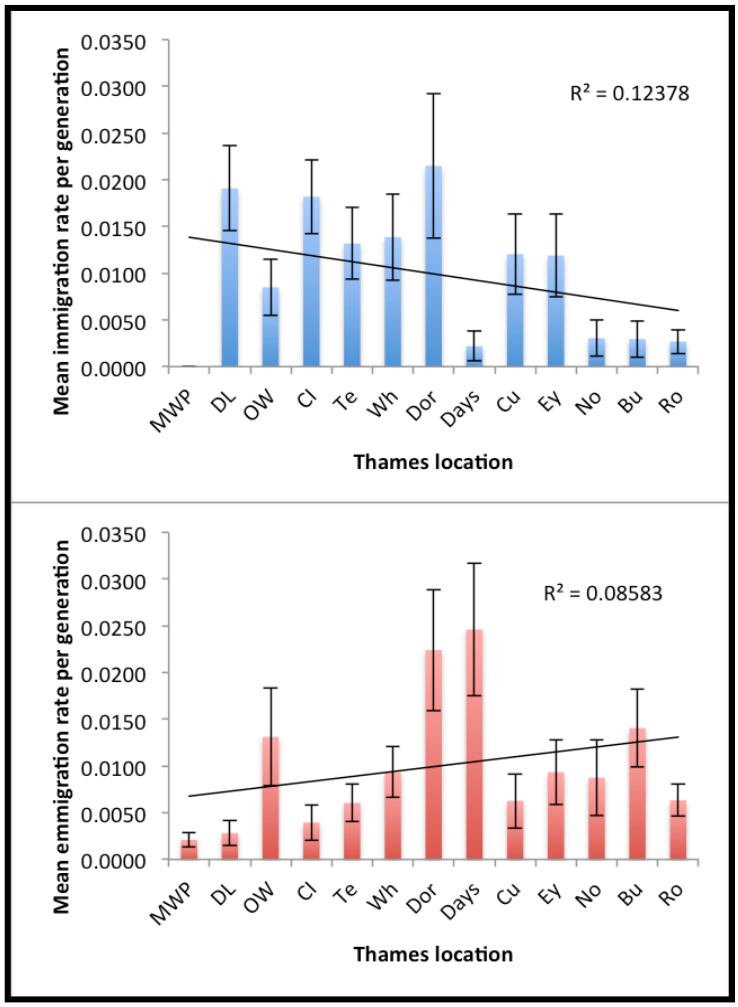


Figure 4.14: Mean values (with standard error bars) of immigration (top) and emigration (bottom) found within Thames populations. Lines of best fit are shown.

The association would be even more positive if Shalford Weir was also removed (Pearson's r = 0.902, 2-tailed p-value = 0.0139). No detectable migration into Shalford Weir was inferred. Unlike the Thames, the trend within the Stour is for the amount of outgoing migrants to decrease with distance upstream, but this is very poorly supported, even if Shalford Weir is excluded (Pearson's r = 0.06, 2-tailed p-value = 0.878 before exclusion, Pearson's r = 0.417, 2-tailed p-value = 0.304). However, a strong upstream correlation would be observed if once again the two most downstream populations were excluded as well (Pearson's r = 0.650, 2-tailed p-value = 0.052).

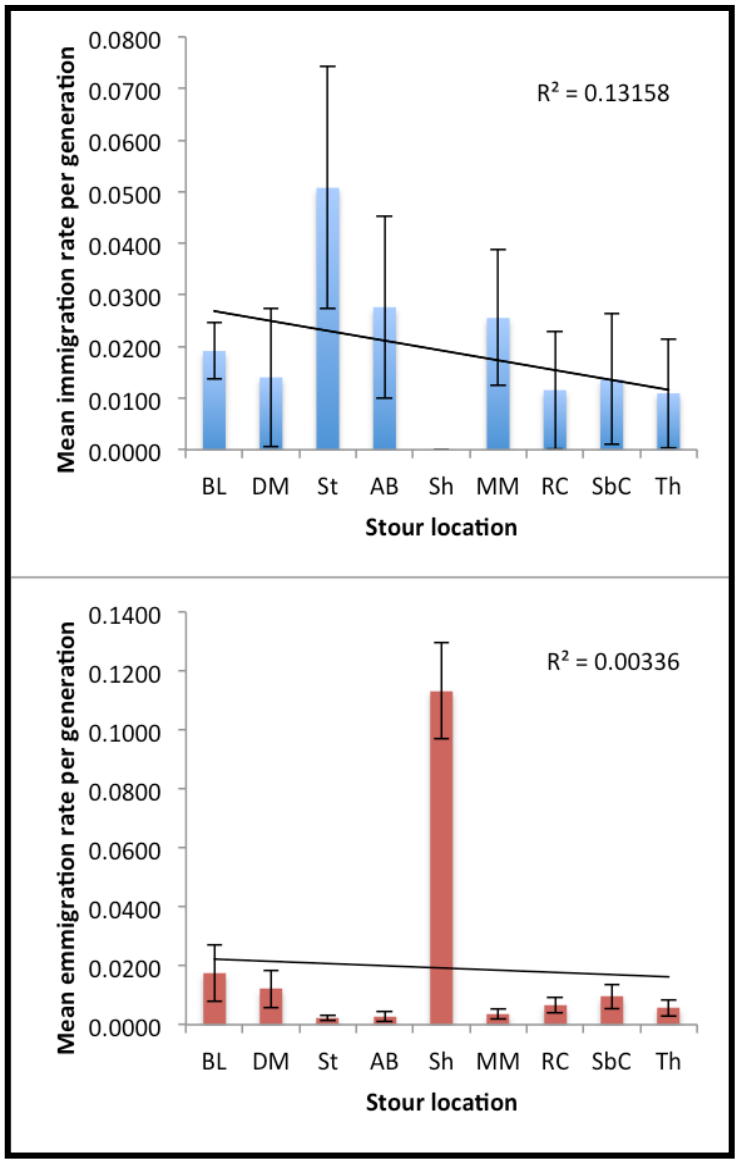


Figure 4.15: Mean values (with standard error bars) of immigration (top) and emigration (bottom) found within Stour populations. Lines of best fit are shown.

As can be observed from Fig 4.16, there is a slight negative correlation between pairwise geographic distances and pairwise rates of exchange of migrants for both rivers, albeit a statistically insignificant one (Thames: Pearson's $r = 0.175$, 2-tailed p-test = 0.128; Stour: Pearson's $r = 0.246$, 2-tailed p-test = 0.148).

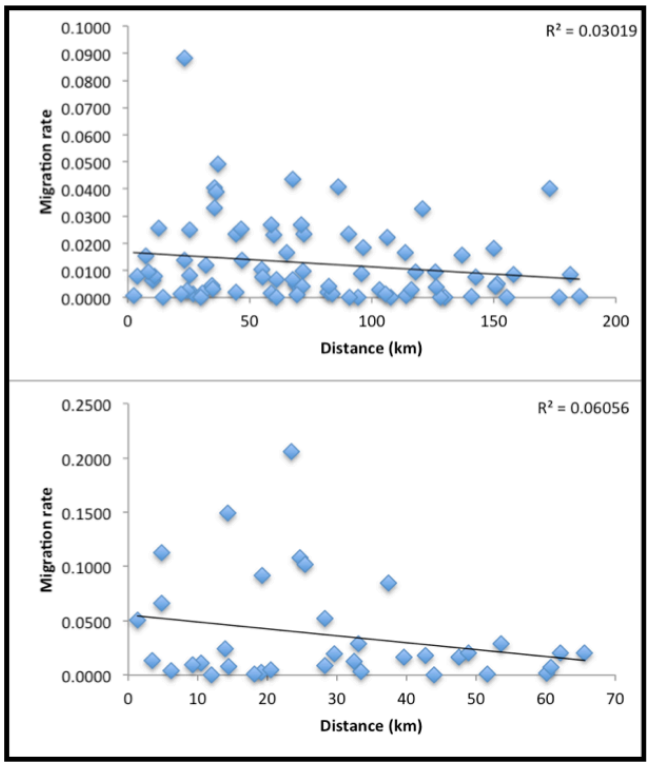


Figure 4.16: Pairwise estimates of migration rate (in either direction) in relation to river distance between sites in the Thames (top) and Stour (bottom), respectively. Lines of best fit are shown.

4.4.4.4 Landscape Genetics

Both rivers displayed a positive relationship between geographical and genetic distances between sampling sites (Fig. 4.17), but this indicated significant isolation by distance by roach populations in the Stour only, but not in the Thames, although the Thames was borderline significant (Thames: Z-statistic = 260.61, $r = 0.185$, $p = 0.087$ and $p = 0.058$ if either but not both distances were log-transformed; Stour: Z-statistic = 59.68, $r = 0.583$, $p = 0.001$, regardless of any log-transformation of distances). The analysis was repeated, this time omitting the pooled populations (Eynsham from the Thames and Stoke by Clare from the Stour). Little change was observed to levels of overall significance: Thames: Z-statistic = 238.93, $r = 0.211$, $p = 0.073$; Stour: Z-statistic = 43.31, $r = 0.676$, $p = 0.002$).

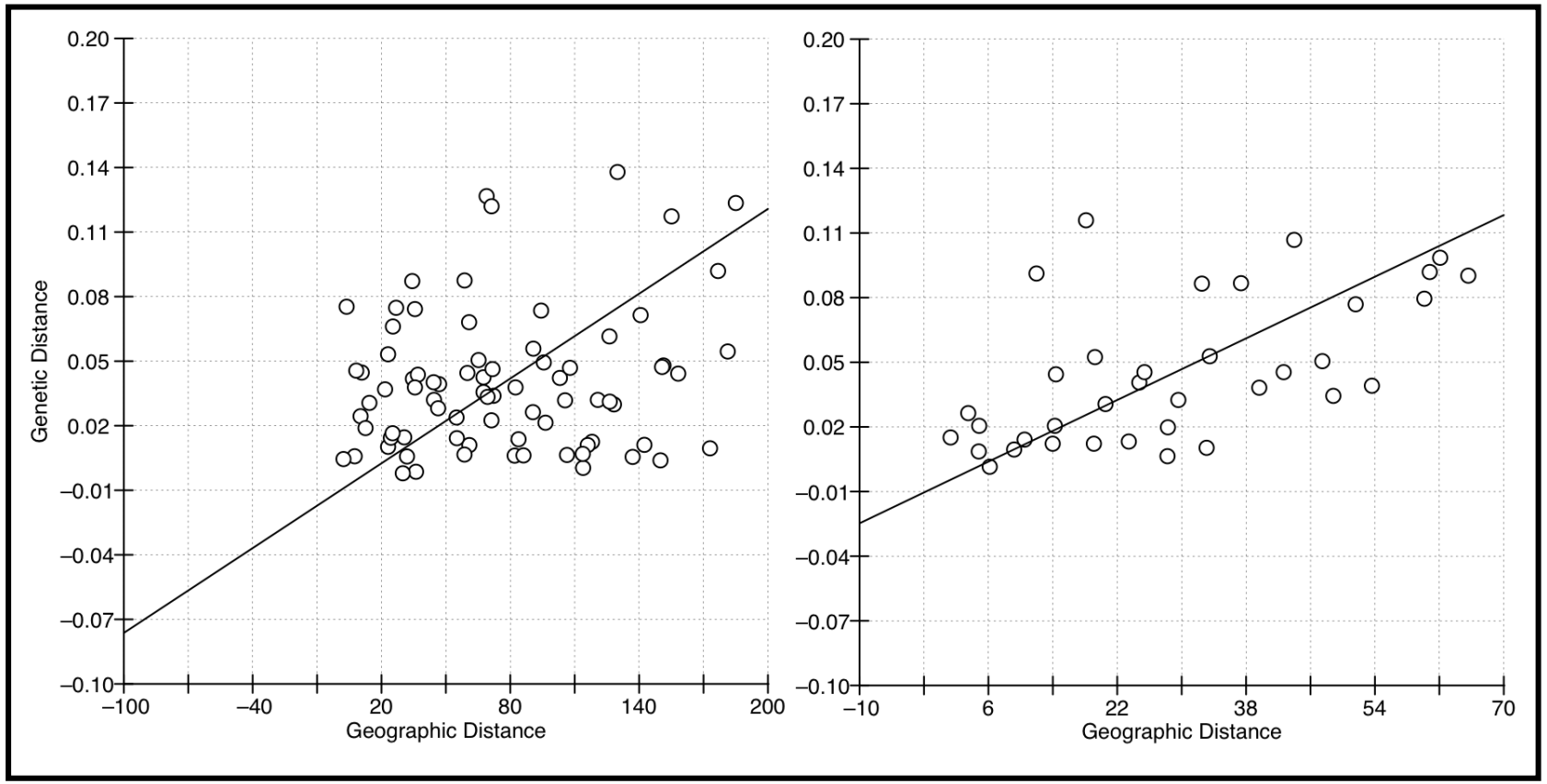


Figure 4.17: Results of the IBD analysis for the Thames (left) and Stour (right), respectively, displaying the graphical relationship between pairwise genetic and geographic distances. Geographic distances are in kilometres (km); genetic distance is described by the equation (θ/(1-θ)). The line of best fit is shown for both graphs.

The results of the genetic autocorrelation analyses are shown in Fig 4.18. The autocorrelations for both the Thames and the Stour showed that the relatedness coefficient, *r*, is significantly higher than expected by chance for both the 5km and 10km distance classes in the Thames (p = 0.001 and 0.010, respectively), and just the 5km distance class in the Stour (p = 0.010). The intercept of each river correlogram at the x-axis was 17.091km and 8.846km, for the Thames and Stour, respectively.

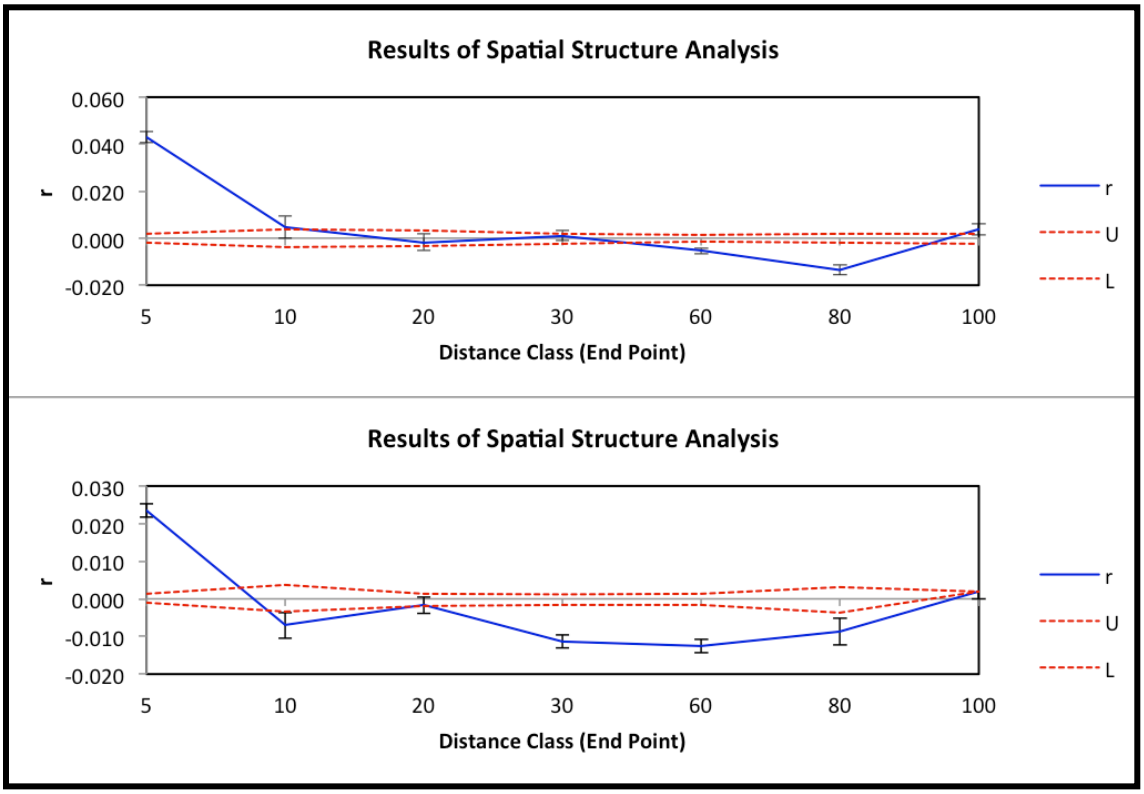


Figure 4.18: Genetic autocorrelation analyses of the Thames (upper panel) and Stour (lower panel) roach populations. Blue line shows the relationship between individual genetic relatedness, *r*, with distance class. The dotted red lines show the upper and lower 95% confidence intervals about the null hypothesis of no difference. The errors bars indicate 95% CI about each point estimate of r, for each distance class.

As with the simple Mantel tests presented in Fig 4.17, the simple and partial Mantel tests – as shown in Table 4.23 – determine the correlation of factors between individual pairwise comparisons. The first block of rows in Table 4.23 show the simple correlation between each of five environmental factors with both genetic diversity (Weir & Cockerham's θ) and recent migration rate.

Table 4.23: Simple and partial Mantel tests to infer correlative processes between the extent of pairwise estimates of genetic differentiation and migration (as previously inferred from BAPs analysis) with five external factors: geographic distance (km), mean differential flow velocity between sites (m^3sec^{-1}), mean differential concentration of dissolved oestradiol between sites (ngL^{-1}) and the number of both weirs and major tributaries (rivers, streams and rivulets) between sites.

		River							
		Stour				Thames			
		Parameter				Parameter			
Variable	Control	θ		Migration rate		θ		Migration rate	
		r	*p*-value	*r*	*p*-value	*r*	*p*-value	*r*	*p*-value
Distance (km)	-	0.536	0.002	-0.246	0.023	0.161	0.121	-0.174	0.066
Mean flow	-	0.121	0.188	-0.102	0.221	0.021	0.411	0.104	0.193
Mean oestradiol	-	0.396	0.013	-0.110	0.229	-0.015	0.560	-0.200	0.066
No. of weirs	-	0.454	0.002	-0.217	0.033	0.235	0.050	-0.182	0.061
No. of tributaries	-	0.693	0.000	-0.287	0.027	0.380	0.023	-0.235	0.032
Distance (D)	F	0.615	0.002	-0.240	0.095	0.183	0.104	-0.289	0.015
	O	0.570	0.002	-0.244	0.026	0.208	0.089	-0.074	0.274
	W	0.436	0.098	-0.147	0.232	-0.170	0.079	-0.011	0.462
	T	-0.302	0.027	0.034	0.422	-0.406	0.012	0.075	0.276
Mean flow (F)	D	-0.375	0.015	0.085	0.348	-0.090	0.310	0.255	0.026
	O	0.569	0.003	-0.239	0.074	0.027	0.380	0.177	0.071
	W	-0.211	0.124	0.036	0.466	-0.221	0.072	0.346	0.007
	T	-0.295	0.036	0.031	0.460	-0.314	0.025	0.341	0.005
Mean Oestradiol (O)	D	0.449	0.000	-0.105	0.264	-0.140	0.221	-0.124	0.177
	F	0.649	0.002	-0.242	0.105	-0.022	0.530	-0.245	0.033
	W	0.415	0.003	-0.102	0.273	-0.170	0.151	-0.124	0.175
	T	0.247	0.045	-0.017	0.453	-0.288	0.036	-0.087	0.262
Number of weirs (W)	D	-0.312	0.154	0.087	0.346	0.251	0.039	-0.055	0.307
	F	0.480	0.007	-0.196	0.115	0.318	0.037	-0.373	0.003
	O	0.474	0.003	-0.213	0.038	0.287	0.036	-0.091	0.230
	T	-0.456	0.008	0.077	0.373	-0.459	0.010	0.145	0.128
Number of tributaries (T)	D	0.581	0.000	-0.155	0.197	0.516	0.018	-0.176	0.095
	F	0.720	0.000	-0.271	0.069	0.478	0.025	-0.395	0.002
	O	0.649	0.000	-0.267	0.014	0.464	0.018	-0.153	0.123
	W	0.694	0.000	-0.206	0.138	0.534	0.016	-0.208	0.064

Genetic differentiation in the River Stour was positively correlated with each potential correlate with the exception of mean flow velocity, before applying partial controls. Of the five variables, only mean difference in oestradiol concentration and the number of tributaries between populations retained significance at the 0.05 alpha level when all other factors were controlled for (distance: $p = 0.002$; oestradiol: $p = 0.0132$; number of weirs: $p = 0.0018$; and number of tributaries: $p < 0.001$). However, if multiple comparisons are taken into account, the correlation of genetic divergence with oestradiol loses significance

when the number of weirs or tributaries is included. The number of weirs was chief amongst the conflationary tertiary variables when applied to the correlation between θ and distance and mean flow rate. A different pattern is observed in the Thames. Of the simple Mantel tests, θ correlates with number of weirs (p= 0.050) and number of tributaries only (p= 0.023). Apart from controlling for the influence of tributary number (p = 0.025), mean levels of oestradiol do not correlate with θ, likewise tributary number with geographic distance and mean flow rate. However, both the number of weirs and tributaries both correlate with θ to a significant extent when controlling for all other factors (although this is not retained under conservative correction).

Recent migration rate is negatively correlated with distance (p = 0.026), number of weirs (p = 0.033) and number of tributaries in the Stour (p = 0.024), although no single factor is overwhelmingly influential. Similarly, in the Thames, only the number of tributaries is correlated with migration in the simple test (p = 0.032), but distance, oestradiol levels and numbers of weirs are borderline significant. Only the correlation of migration rate with mean flow gains and retains significance when all other factors are incorporated (with the exception of mean oestradiol (p = 0.071).

In addition to the pairwise analysis of genetic divergence and connectivity, one sought to find any correlation between site-specific factors and levels of genetic diversity at specific sites along both the Stour and the Thames, and across all sites inclusive. All indices of genetic variation listed in Tables 4.7 and 4.8 were assessed for co-variance with mean oestradiol and flow rates per site, as well as geographic distance and mean snout-vent length of the individual fish (a proxy of age-class distribution) (see Table 4.24 for details). Very little patterning is evident among population means and measures of diversity. There is a positive correlation between the mean length of roach within a population and the degree of population-level inbreeding (F_{IS}) and a concordant negative correlation in observed heterozygosity with increased size in the Stour, which attains borderline significance if the Shalford Weir population is removed (which has the lowest F_{IS} and highest H_O scores of 0.034 (p = 0.057) and 0.649 (p= 0.063), respectively). The Stour also exhibits a weak negative correlation between the mean number of alleles per locus and distance upstream (p = 0.071). However, performing multivariate regressions of all extraneous factors against the five measures of genetic diversity does not result in any single factor being statistically important within rivers or across all populations, although sample size is a significant predictor of MNA in the Thames alone.

Table 4.24: Mean levels of environmental factors for each population in the Thames and the Stour: Distance = km; mean length = cm; mean oestradiol = ng L^{-1}; and mean flow = m^3 sec^{-1}.

River	Distance Upstream	Sample size	Mean length	Mean oestradiol	Mean Flow
Thames					
MWP	0	33	121.710	2.521	42.519
DL	3.755	65	133.540	2.664	48.060
OW	26.82	49	145.300	2.591	38.545
CI	34.33	63	131.270	2.285	46.477
Te	58.72	60	155.250	2.161	45.933
Wh	68.98	23	132.080	1.882	29.081
Do	71.27	33	123.620	2.465	4.138
Da	94.32	33	132.890	2.181	23.257
Cu	129.92	40	156.380	1.988	23.670
Ey	140.72	33	139.160	1.392	12.593
No	155.1	26	108.650	1.392	12.593
Bu	176.78	27	125.470	1.806	9.189
Ro	185.04	22	127.570	0.431	1.536
Stour					
Br	0.000	82	92.530	1.887	3.178
DM	13.3	34	97.180	1.866	3.186
St	18.33	59	62.690	1.830	3.097
AB	26.33	71	126.220	2.050	2.107
Sh	32.71	34	90.010	2.530	1.972
MM	37.56	63	67.940	14.826	0.007
RC	61.49	17	107.630	2.606	0.853
SbC	62.1	37	138.290	3.227	0.730
Th	65.6	20	103.500	0.691	0.251

Geneland identified three distinct, geographically defined populations (Fig 4.19). The first population consists of the most upstream Stour population –Thurlow - with all populations of the Thames, with the exception of the most downstream site, MWP. The second population consists of the remaining Stour populations; and MWP is accorded distinct population status of its own.

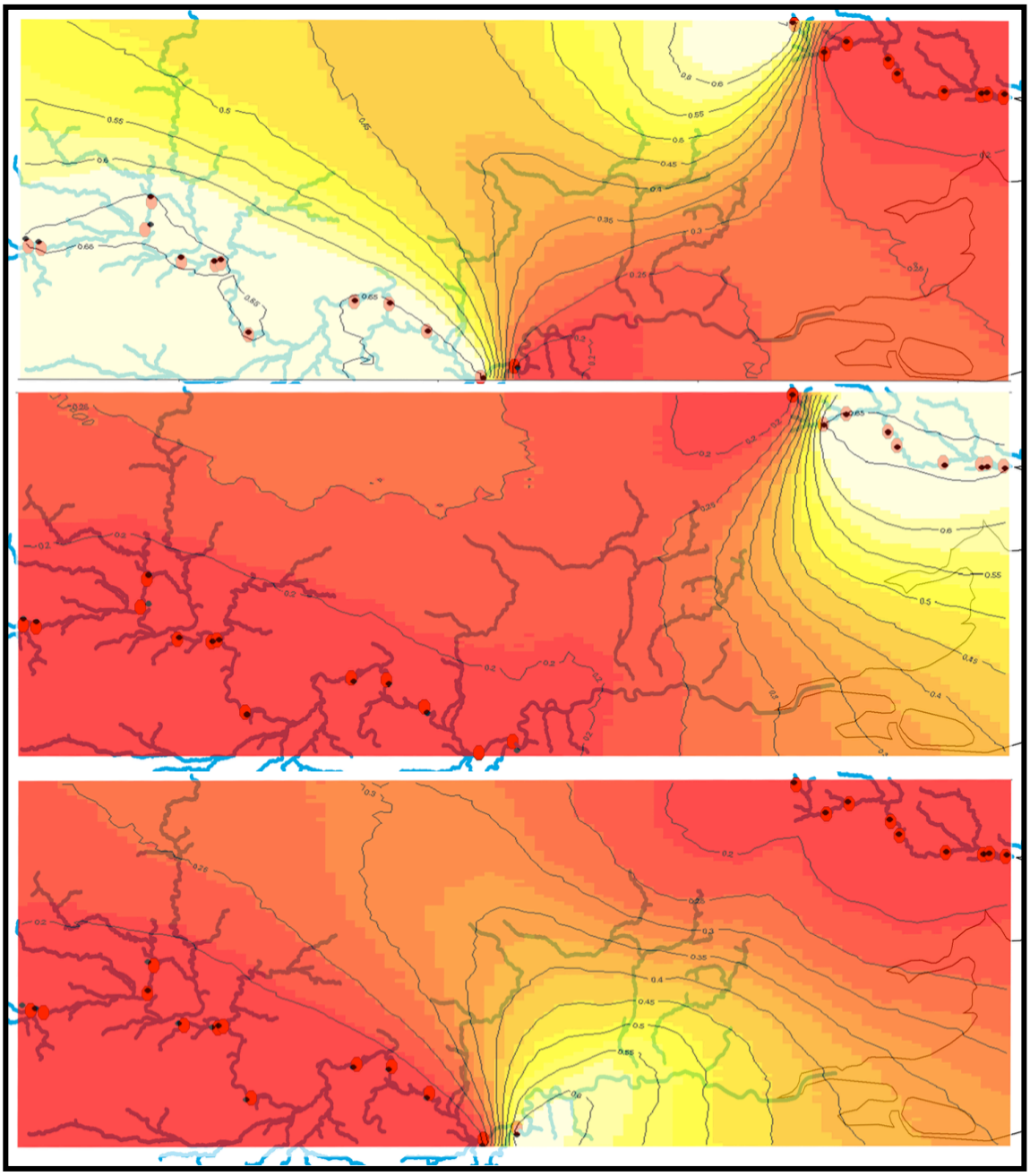
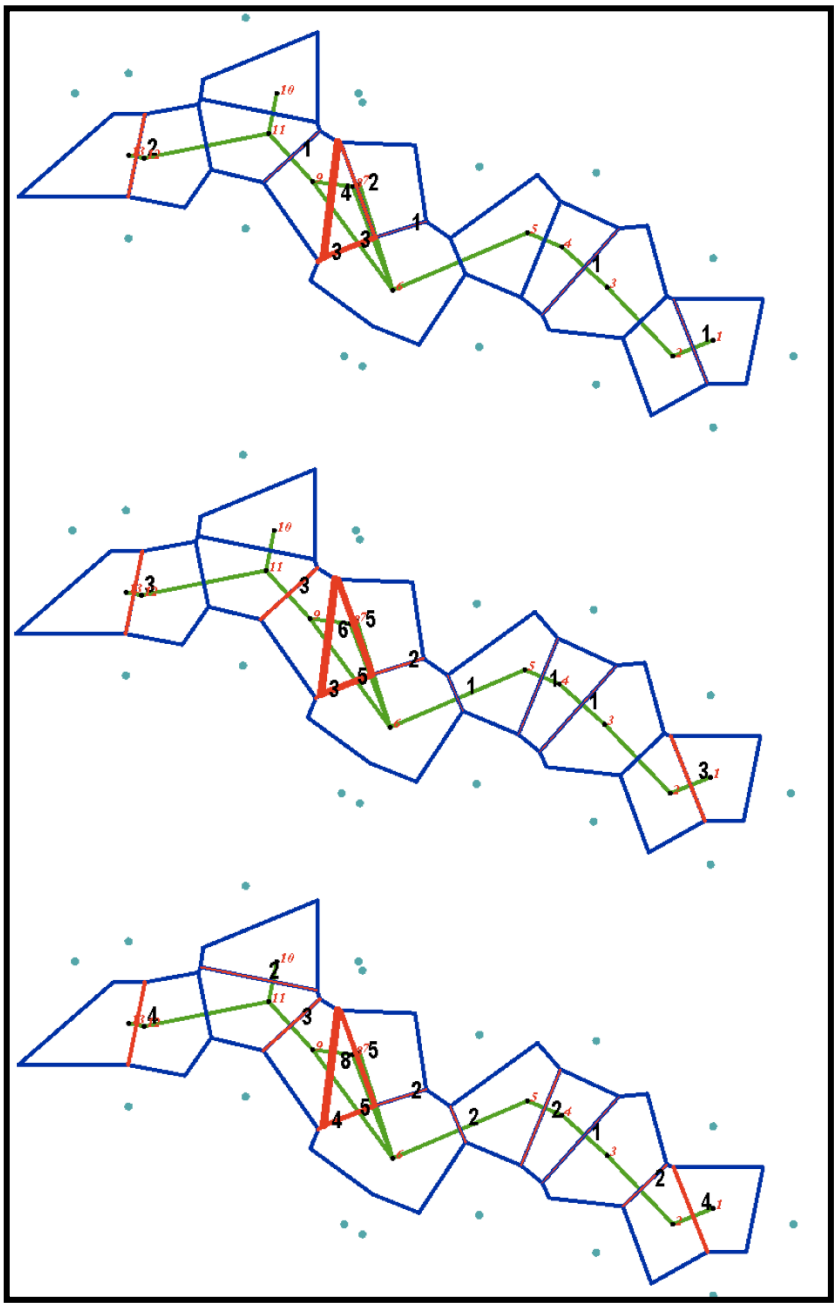


Figure 4.19: Geneland analysis. Posterior probability contours for K = 1 – 3 (top to bottom) showing the probability with which the 22 sampled sites across both rivers belong to each of K = 3 populations. Low probability contours are indicated by red, increasing in likelihood with increasing brightness and lightness.

The final analysis utilized Monmonier's algorithm to identify spatially identifiable genetic breaks (Figs 4.20 and 4.21).



.

Figure 4.20: Graphical representation of the Thames (thin green line calculated by Delaunay triangulation) and superimposed inferred genetic breaks (thick red lines). Top diagram displays the thickness of the barrier as a proportion of the ten microsatellite loci that identifies this break as being the most identifiable. Middle diagram incorporates the second choice and the bottom diagram incorporates the third choice. The number of supportive loci is shown (bold type)

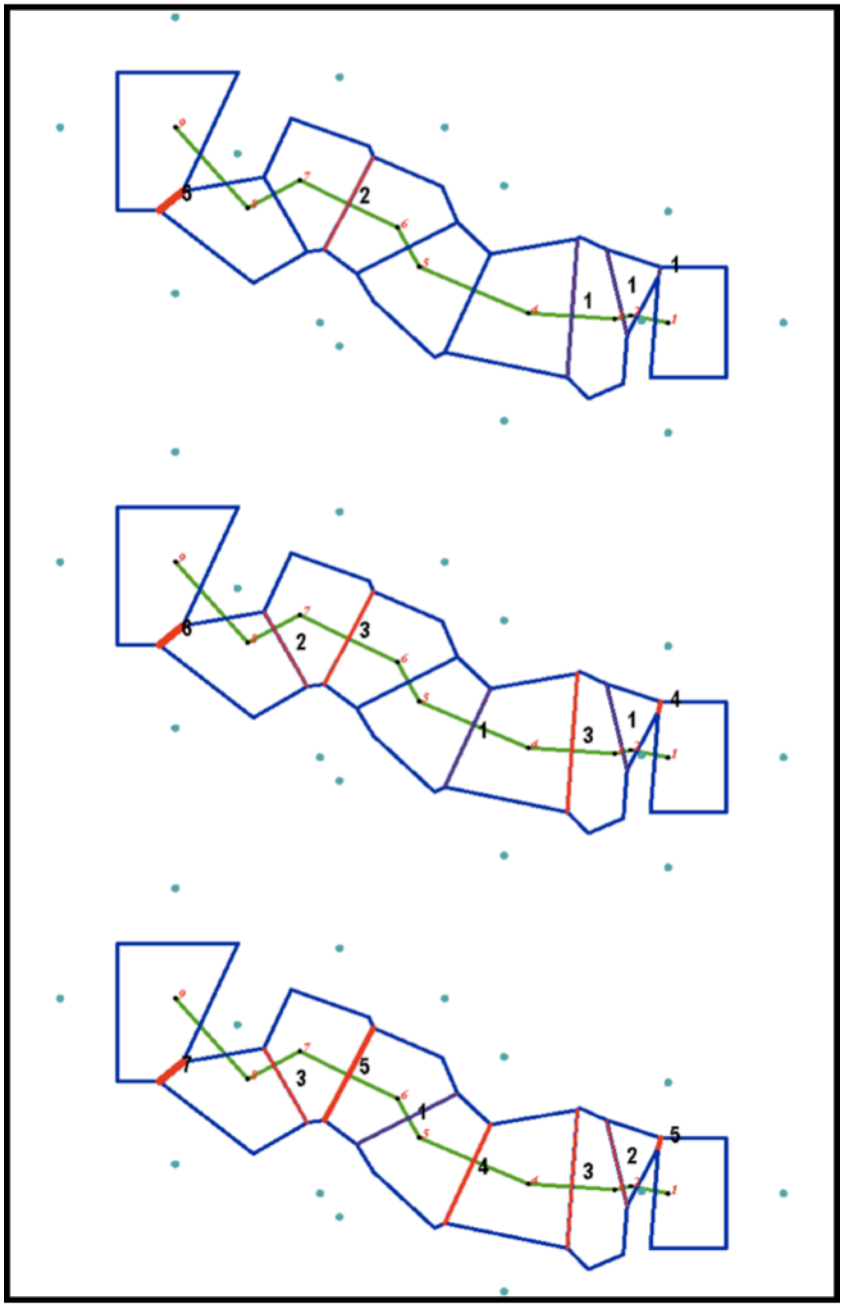


Figure 4.21: Graphical representation of the Stour (thin green line calculated by Delaunay triangulation) and superimposed inferred genetic breaks (thick red lines). Top diagram displays the thickness of the barrier as a proportion of the ten microsatellite loci that identifies this break as being the most identifiable. Middle diagram incorporates the second choice and the bottom diagram incorporates the third choice. The number of supportive loci is shown (bold type).

The most significant genetic break identified within the Thames is between the population at Culham and downstream populations (Days and Dorchester, given the peculiarities of the triangulation procedure) (4/10 loci support a break between Culham and Days, top panel Fig 4.20). Incorporating the second and third choices, the populations upstream of the confluence of the Thame and Thames (Days and Dorchester) are separated from the immediate downstream site at Whitchurch (5/10 loci, bottom panel) and the upstream site

of Roundhouse is deemed to be under the influence of a break from the downstream site at Buscot (4/10 loci, bottom panel), in addition to MWP being distinct with the same number of loci. The space between Culham and Eynsham is also inferred to experience some break in genetic continuity (3/10 loci).

The evidence for genetic breaks in the Stour involves the apparent distinction of the Thurlow population (5/10 loci, first choice, top panel, Fig 4.21). However, incorporating second and third inferences, significant genetic breaks are identified between Brantham Lock and upstream (5/10 loci), Mill Meadow and upstream (5/10 loci), and Anchor Bridge and Shalford Weir (4/10 loci).

4.4.5 Discussion

4.4.5.1 Departure From Migration-Drift Equilibrium.

Population contractions can leave a significant imprint upon levels of expected heterozygosity relative to that expected under mutation-drift equilibrium conditions. None of the Thames populations showed any sign of a recent population contraction, at least in the last 0.2 – $4N_e$ generations. In contrast, the Stour was inferred to have undergone population contractions at four locales: Brantham Lock, Mill Meadow, Stoke-by-Clare and Thurlow. The latter two populations are found upstream in the most lotic part of the Stour. Such upstream populations are more likely to undergo demographic contractions due to inherently smaller local population sizes (smaller habitats), the greater difficulty faced by migrants travelling to upstream sites and the increasingly patchy nature of spawning habitat with increasing distance upstream. The inferred demographic reductions at Mill Meadow and Brantham Lock are more difficult to explain on the basis of population connectivity and population size. Here, one may refer to the potential impact of environmental factors. Brantham lock is within the tidal range of the Stour and roach here must display tolerance to an increase in dissolved salt, either physiologically or behaviourally. At Mill Meadow, the mean concentration of dissolved oestradiols – known to feminize male roach – are extremely high (over 14ng L^{-1}), more than five-fold greater than at any other site sampled in the river (but see below). It is possible that the action of selection may have had an influence, for example through the action of selective sweeps during periods of high stress in environments such as these, but the impact of directional selection on BOTTLENECK results is not well understood (Luikart et al. 1998). The lack

of evidence for bottlenecks in the Thames – with its long history of anthropogenic interference – could be the result of larger population sizes within this river, that enough time has elapsed since a bottleneck for any genetic signal to have eroded, or due to the influence of false negative results. The reliability of Piry et al.'s software in determining actual instances of bottlenecks by assessing excessive levels of heterozygosity under neutral expectations has been called into question when tested on known bottlenecked versus known non-bottlenecked populations (Cristescu et al. 2010). Additionally, the signature of a bottleneck is strongly linked to the severity and duration of such an event and does not always result in the loss of significant levels of genetic variation (e.g. Demandt and Björklund 2007), as well as sampling scheme, population structure and overall levels of genetic diversity (Chikhi et al. 2010).

The value of identifying populations that have undergone demographic contractions is important. One may identify populations that naturally have to be replenished by individuals migrating from elsewhere, and thus lines of genetic exchange between these populations need to be kept open. Almost all populations in this study were found to both donate and receive migrants (see below). However, if natural sources of migrants are blocked, then the chances that such populations become extinct are increased (Lande 1994; Lynch et al. 1995). Bottlenecks, inferred from nDNA, have been identified across freshwater and diadromous freshwater fishes using BOTTLENECK (e.g. bullhead (Hänfling and Weetman 2006), dace (Costedoat et al. 2006), the Asiatic cyprinid *Labeo dero* (Chaturvedi et al. 2011) and grayling (Dawnay et al. 2011)) but not in other species in which one would expect population contractions to have occurred, but have been identified using other methods of inference such as MSVAR (Beaumont 1999) (e.g. the endangered Iberian cyprinid *Chondrostoma lusitanicum* (Sousa et al. 2008)). Hänfling et al. (2004) did not detect any signs of population bottlenecking in isolated riverine or floodplain populations of roach based upon allozyme data, with high effective population sizes and likely genetic replenishment by periodic flooding given as the primary reason, despite the data being sensitive enough to detect bottlenecking.

4.4.5.2 Population Sub-structuring within the Thames and Stour.

Both rivers displayed significant population subdivision, albeit only on a low to moderate scale (mean global F_{ST}: Thames = 0.032; Stour = 0.039). Levels of population differentiation were higher within the Stour than within the Thames, but not significantly

so. Levels of differentiation, as measured by F_{ST}, are of a similar if lower level than that found in allozyme studies of roach (e.g. Laroche et al. 1999; Hänfling et al. 2004), a result consistent with high within-population diversities. Microsatellite loci tend to possess more alleles that may be shared between populations, a result that highlights that the application of such loci may be slightly problematic in organisms expected to be highly genetically diverse and which share recent geological histories.

Levels of differentiation among populations within both the Thames and the Stour (and globally) are low to moderate in extent (values less than 0.05 are deemed to be low by some authors (Balloux & Lugon-Moulin 2002)). For example, long isolated populations of the arctic charr from two lakes in Scotland possess non-standardized estimates of F_{ST} that span 0.091 – 0.439, including sympatric populations (Adams et al. 2008). Similarly, endangered Iberian cyprinids such as *Chondrostoma lusitanicum* display levels of differentiation (average F_{ST} = 0.390) an order of magnitude greater than any comparable roach example (Sousa et al. 2008). However, the global F_{ST} of 0.020 found between the pooled datasets of the Thames and Stour was similar to the value found for populations of the tench between two lake sites in Germany (Kohlmann et al. 2007). Similarly, within a Swedish lake, samples of perch displayed an F_{ST} of 0.023 (range 0.001 – 0.040). Chub of the River Durance, France, displayed significant differentiation in 37% of pairwise comparisons, with an F_{ST} range of 0 – 0.055 (Dehais et al. 2010). By contrast, the proportion of significantly different pairwise comparisons in the roach was 75.6% in the Thames (F_{ST} = 0 - 0.109) and 88.9% in the Stour (F_{ST} = 0 – 0.101). Dawnay et al. (2011) found pairwise comparisons of θ significant for all but one grayling population sampled within the UK. All but one of pairwise comparisons of F_{ST} of 26 sampled locations across Europe was significant for divergence in the bitterling (Bryja et al. 2010), which is expected in rheophilic species that inhabit faster-flowing streams. Generally, the range of genetic divergence found in the roach may be located within the higher end of the spectrum of genetic divergences found in most microsatellite studies of European freshwater fishes.

Aside from classic and standardised estimates of genetic structuring, the present study also utilised spatial data and individual assignment to determine the number (K) of hypothesised populations indicated by the data, whose boundaries are not obvious from an external observation of individuals or environment. The STRUCTURE analysis identified two and five HWE populations, dependent upon whether one inputted specific location data into the analysis. The first analysis confirmed that the Stour and the Thames

individuals generally agree with their location of sampling, i.e. that the Stour and Thames may each be considered as genetically cohesive units. This finding is supported by both the PCA and individual assignment analyses. Using sampling location, the Thames population of MWP and the Stour population of Thurlow were recognised as being genetically distinct from other populations within their respective rivers, supporting the theta results. Additionally, when either the Thames or the Stour were analysed using STRUCTURE separately, K = 2 was deemed most likely for both rivers, regardless of whether sampling location was inputted as a prior, with MWP and Thurlow again identified as being distinct populations from the remainder of the roaches in either river (see Appendix B). However, the Geneland analysis supported only K = 3 populations, such that Thurlow was grouped in with the non-MWP Thames populations.

A number of studies have conducted similar spatial *post hoc* analyses of genetic clustering in European cyprinid species. Sousa et al. (2008) found that rivers within the Samarra drainage of Portugal possessed their own unique signature of genetic variation within the endangered cyprinid *Chondrostoma lusitanicum*, as described by a STRUCTURE analysis, but with no differentiation in the smaller southern drainage of Sines. At the other extreme, the widely distributed bitterling possesses much less structuring, with K = 2 or 3 populations best describing the distribution of genetically cohesive clusters from 26 localities across Europe. The bitterling clusters inferred coincide with those that once were connected to the Danube, a Dniester-Dnieper-Vistula cluster (including an English population) and Aegean / Asia-Minor cluster (which groups with the Danube cluster when K = 2) (Bryja et al. 2010). The vairone of the Italian peninsula were inferred to possess 5 homogeneous groups, which correspond to either individual rivers or small assemblages of river populations (Marchetto et al. 2010). Similarly, 5 population clusters were reported from a UK survey of the grayling, roughly coinciding with Scottish, northern, midland and southern populations, plus a cluster composed of populations widely distributed within the UK (Dawnay et al. 2011). Demandt (2010) found no evidence in roach of multiple HWE populations in the Forsmark Biotest station lake, Sweden, but K = 4 in the European perch. The chub of the River Durance also failed to yield any significant genetically differentiated groups (Dehais et al. 2010). This study suggests that nuclear genetic clustering into definable populations can be made on a finer geographical scale in roach than that observed in most other European cyprinids.

Physical population fragmentation has been found to significantly impact upon measures of genetic diversification in cyprinid species (Blanchet et al. 2010). In chub and dace, levels of F_{ST} were significantly higher in the fragmented River Viaur compared to the continuous River Célé[39], a pattern not found in two smaller species, gudgeon *Gobio gobio* and minnow *Phoxinus phoxinus* (Blanchet et al. 2010). As well as F_{ST}, Blanchet et al. (2010) found significant reductions in mean allelic richness and mean observed heterozygosity between the fragmented environment and the continuous one in three of the four species (the minnow excepted). This is an observation that is seen with levels of allelic richness ($p < 0.001$) and observed heterozygosity ($p = 0.0194$) detected between the Thames and Stour roach, with levels of mean allelic and observed heterozygosity significantly less in the Stour than in the Thames. Blanchet et al. (2010) also found no difference in effective population sizes between the two rivers.

Blanchet et al. (2010) make little reference to the quality of feeding and spawning habitat in their study, instead defining a continuous landscape as a riverine environment with few anthropogenic blockages across the stream/river passage, regardless of whether these sometimes ancient mills/weirs are still functional or otherwise. By this definition, the Stour would be classified as the more fragmented landscape, and the results found in this study (lower allelic richness, observed heterozygosity and greater overall level of structuring, F_{ST}) are consistent with the findings of Blanchet et al. (2010), particularly with regard to the larger and more vagile species (chub and dace). Unlike the species studied in Blanchet et al. (2010), the roach of the Thames and the Stour exhibited significant variation in size and in mean distribution of size classes along these rivers ($p << 0.001$, heteroscedastic student's t-test).

4.4.5.3 Spatial Patterning of Genetic Diversity

In unidimensional river environments, one may expect that the physical environment would place greater emphasis on downstream gene flow, potentially eroding historical sources of genetic diversity (in upstream locations) with source-sink metapopulation dynamics, which are further complicated by contemporaneous anthropogenic barriers to gene flow (Weetman and Hänfling 2006; Blanchet et al. 2010). However, the continual

[39] Fragmented here refers to the presence of anthropogenic weirs and mills. More than 50 are to be found along the main channel of the River Viaur (on average one per 2-3km), most of which originate from the 15th century with few fish-passes erected. Additionally, two dams were built 30km and 80km from the source, respectively. By contrast, the River Célé posseses only ten weirs, each built during the 20th century and equipped with fish ladders (Blanchet et al. 2010).

existence of upstream populations despite pressures to redistribute organisms downstream (the so-called 'drift paradox', Müller 1954) requires explanation. One resolution for freshwater fish species involves either the continued replenishment of individuals at a rate equal to the rate at which larvae are lost (Anholt 1995) (in which case high levels of genetic variation may be maintained), or upstream areas are recolonised after local extirpations (in which case one would expect low levels of within-population genetic diversity, consistent with founder event dynamics). A general expectation is that levels of genetic diversity would decrease with distance upstream (Hänfling & Weetman 2006; Dehais et al. 2010) because of less available suitable habitat, smaller population sizes (due to lower water volume and increased competition), and through the loss of alleles via downstream juvenile drift; or a combination of such factors. This study shows that in both rivers, roach are able to actively migrate and transfer their genes to populations that are located upstream as well as downstream, such that overall levels of genetic diversity are fairly high, even in the lotic headwaters of the Stour. However, in the Stour in particular, one observed an important distance component to the distribution of genetic diversity suggesting locally important equilibrium conditions, which become more perturbed with upstream location (e.g. the two most upstream populations may represent sink populations, leading to their relative genetic isolation). This is exemplified by the distribution of effective sizes within both rivers, calculated according to Jones & Wang (2009). The effective sizes of Stour populations show a strong negative correlation with distance upstream ($r = 0.657$, 2-tailed p-value = 0.055), whereas the Thames displays a similar correlation if MWP is omitted ($r = 0.679$, 2-tailed p-value = 0.011; if MWP retained: $r = 0.393$, 2-tailed p-value = 0.184) (see Appendix B).

If evolutionary equilibrium conditions are assumed, in a one-dimensional environment with a bi-directional opportunity for effective migration, the patterning of genetic differentiation should show a significant correlation with distance, whereby genetic exchange occurs at a greater frequency between populations when the distance between them decreases (Kimura and Weiss 1964). This stepping-stone model would be disrupted if significant long-distance migration occurred, with an emphasis in riverine environments of significant movement of larvae and fry downstream and / or long-distance adult movements upstream. Freshwater fish have been found to display significant IBD patterns in genetic diversity, for example all *Telestes* species show highly significant IBD within the Rhône (Dubut et al. 2012). In the present study, roach populations displayed a significant correlation between genetic differentiation and distance over some 66km of river within the Stour ($p = 0.001$) and borderline significance over 185km of the Thames

(p = 0.087). The pattern of IBD in the Stour is evocative of Hutchinson and Templeton's (1999) model of regionalised equilibrium between genetic drift and migration, whereas the Thames reflects a model of IBD whereby inter-population migration is higher or where long-distance migration is a more important factor than it is in the Stour. The Stour model of IBD is similar to that found in the bullhead (*Cottus gobio*) along the River Seven, UK (Hänfling and Weetman 2006), and in intermediate sized cyprinids such as the dace and gudgeon in the anthropogenically disrupted river Viaur, France (Blanchard et al. 2010). Similarly, the intermediate sized species in the Blanchard et al. (2010) study show levels of reduced IBD as observed here in the Thames roach, where fewer human-built occlusions exist and where mean roach size is higher.

When isolation by distance occurs within unstable environments within metapopulations, population equilibria become dependent upon migration, extinction rates *and* the founder effect, in addition to drift (Levins 1970). The primary prediction of propagule-colonization models of equilibrium dynamics (Slatkin 1977; Maruyama and Kimura 1980) is that they predict less retention of diversity across the entire metapopulation, perhaps explaining the greater within-population structuring and lower allelic richness and observed heterozygosity found in the Stour relative to the Thames where a metapopulation dynamic less evident. Wright predicted an increase in F_{ST} in an early invocation of founder effects relating to his island models (1940). Even if local extinction is common, the extinction rate seems to have little effect upon overall levels of F_{ST} (Wade and McCauley 1988), so long as there remains a source of colonization. Given the roach's ability to disperse long distances, then one would predict that the Stour population in particular to exhibit similarities to these modified population models.

The presence of IBD or any level of structuring in freshwater fishes is dependent upon the availability of suitable spawning habitat. In closed lentic systems, roach are observed to have a high degree of natal spawning philopatry, exhibiting such a degree of return migration accuracy that they may better the abilities of salmonids within rivers (e.g. 83.5 – 92.0% in Swedish roach (L'Abée-Lund & Vøllestad 1985) compared with 50% returning Atlantic salmon between 1986-88 in the River Dee, Aberdeenshire (Youngson et al. 1994)). Therefore, the potential for roach natal philopatry on such small scales to effect population structuring is significant. In roach within the Stour, the average distance at which roach relatedness is offset by the action of genetic drift is 8.846km: the resulting effect is a strong IBD signal within this river. Contrasting roach with the more familiar homing abilities of salmonids, the Atlantic salmon of the Varguza River in Finland

(Primmer et al. 2006) exhibited less of an IBD relationship (r_{xy} = 0.33, p = 0.05) and a much larger intercept between r and distance class (intercept = 34km) than in the Stour (r_{xy} = 0.583, p = 0.001, intercept distance = 8.846km), although the Finnish salmon did exhibit greater IBD than Thames roach. Bull salmon *Salvelinus confluentus* of the Kootenay River in British Columbia show much greater levels of IBD (r_{xy} =0.47, $p < 0.001$) than Stour roach, but not in the Pine River (r_{xy} = 0.15, p = 0.147) (Costello et al. 2003). A pattern of IBD is common in freshwater fishes, but not always evident (e.g. bullhead (Hänfling and Weetman 2006); chub (Dehais et al. 2010). pikeperch *Sander lucioperca* (Björklund et al. 2007)). For allozyme data, Hänfling et al. (2004) and Laroche et al. (1999) found no correlation between genetic and geographic distances among roach populations within the Elbe and Rhône Rivers, respectively. It is not wise to always assume IBD, even within species with significant homing abilities; rather individual river populations should be assessed on a case-by-case basis.

4.4.5.4 Population Connectivity

Body length in animals is positively correlated with an ability to disperse greater distances within species (Jenkins et al. 2007). Additionally, within-species population density is inversely related to body size (Cotgreave 1993). Sampling density – a predictor of population density, albeit a poor one - was greatest in the Stour, and these roach also possessed significantly smaller body sizes. The data in this study support the theoretical assertion that dispersal distances – as expressed by the autocorrelation coefficient, r, - would be shorter in the Stour roach population (because they are smaller individuals), where the maximum dispersal distance is inferred to be 50% less than that of roach in the Thames. These data are also consistent with the greater strength of signal of IBD in the Stour than the Thames.

Fig. 4.14 shows the mean estimate of recent migration rate in to and out of each population (sampling area) in the Thames. Generally, the further upstream a population is the more likely it is to yield migrants and less likely to receive them, although the relationship is far from perfect. MWP, as an area likely to harbour a significant number of introduced fish from previous years, receives no migrants, despite being the most downstream located population. That the EA stocked the nearby river with 30000 fish only a few years before, suggests that whilst an introduced population may persist and retain high levels of diversity, expected given the numbers of introduced stock, it retains an aberrant genetic

signal that is incongruous with the genetic signal of the Thames in general. Additionally, the MWP fish appear to contribute little gene flow to other locations within the Thames, even those located proximately. However, by persisting, the MWP does contribute to the spatial genetic diversity of the Thames and of the species.

The number of weirs/mills/locks seemed to impact little upon the migration rate in either river, but some association with patterns of genetic divergence exists in the Stour. The Thames possesses many more fish passes designed to allow the upstream migration of anadromous, catadromous and migratory potamodromous fishes. These passages are effective for pelagic and bentho-pelagic coarse fish, and roach in particular, but less so for more demersal species (Knaepkens et al. 2006). The presence of weirs affects local genetic equilibrium conditions over time that exacerbates the extent of genetic isolation with distance in the demersal piscifauna (e.g. bullhead (Hänfling and Weetman 2006)). In roach, there is a hint that weirs may limit long-term gene flow (e.g. in the Stour), but this situation will have rapidly changed in the last five years or so with the construction of fish passes, particularly at Bures Mill (Environment Agency 2010).

Regardless of the extent of population connectivity among populations in both study rivers, there is evidence for limited gene flow between almost all studied populations. However, this level of connectivity is not sufficient to genetically homogenize roach populations in either river. More migration was detected in the Stour, probably due to it being a much smaller river with more means of facilitating gene flow among spawning areas (greater habitat provision, etc), although genetic structuring was more pronounced in the Stour, MWP aside. This discrepancy may be explained by a high recent migration rate, but historically lower migration rate; and/or by source-sink metapopulation dynamics. The locks, weirs and mills that were once prevalent along its banks and crossings, but are now mostly dis-used or modified for fish passage, may have obstructed historical gene flow. Even so, major obstacles to gene flow and current migration, such as dams, remain a more serious cause for concern as regards affecting the distribution of neutral genetic variation in larger eurytopic and rheophilic cyprinids (Laroche et al. 1999; Dehais et al. 2010).

4.4.5.5 Management Implications

The analysis of migration suggests that some areas of river are more important than others in producing migrants that help populate areas both upstream and downstream. Such a pattern is indicative of a source-sink metapopulation. This pattern was most in evidence within the Stour, whereby over 65% of the immigrant genotypes were derived from the Shalford Weir population, but was also observed in the Days-Culham area of the Thames. The management implications are obvious, if recent immigration rates are reflective of historical migration patterns: emphasis should be placed on conserving and preserving these areas as providers of genetic diversity in perpetuity.

Within the roach populations of these two rivers, there is an apparent paradox between the extent of migration and the degree to which populations are genetically differentiated. Even though recent migration rate is similar within the Thames and Stour, there is more structuring within the Stour than in the Thames. The genetic neighbourhood size is also two-fold smaller in the Stour. The answer probably lies in the fact that at least some of the Stour populations have experienced population crashes, which has led to an increase in the degree of familial inbreeding in these populations, contracting the average neighbourhood size. Brantham Lock is located in the tidal Stour and as such roach here must acclimate to the conditions by assuming, either through phenotypic plasticity or natural selection, a halophilic tolerance. Mill Meadow experiences the lowest flow rate in the Stour, but also the highest concentrations of dissolved oestradiol in both the Thames and the Stour, which may attain toxic levels. Thurlow and Stoke-by-Clare are located either side of the extraction point for the removal of water to feed Essex reservoirs. Both these latter populations are also located in the most upstream and environmentally heterogeneous habitats, which probably support fewer individuals and receive fewer immigrants.

Although precise data are absent, the Thames has a long history of pollution, localised fish kills and recurrent stocking by the Environment Agency. An interesting finding of this study is the low diversity and aberrant genetic signal of the roach at MWP, a site where 30000 individuals were released just downstream of this site six years prior to sampling (15000 at Teddington (geographical co-ordinates: 51.470042 latitude; -0.321241 longitude) and 15000 at London Apprentice, further downstream (geographical co-ordinates: 51.432433 latitude; -0.326072 longitude) (Nigel Hewlett, pers comm)). The migration data suggest that roach from elsewhere have not integrated into MWP, and that roach from MWP make for poor migrants into other areas within the Thames. Whilst the

MWP population has persisted, at least over four years, the time frame may be too short to accurately assess the success of this introduction and a further survey may be necessary. However, the MWP population has the lowest effective population size of any population in the Thames (N = 34, see Appendix B)[40], consistent with a reduced number of breeding individuals in a fish farm environment (i.e. stocked fish). Reduced genetic fitness is associated with population inbreeding, especially in artificial environments such as fish farms, which may account for the poor rate of introgression of MWP migrants into all other populations.

One final implication of this study is that the level of dissolved oestradiol is correlated with the degree to which populations are differentiated from one another within the Stour, although mean levels of oestradiol are not correlated with mean levels of genetic diversity along the length of the Stour (see Fig 4.22 for a map of oestradiol concentrations within the Stour). However, most loci show no relationship between the difference in mean dissolved oestradiol levels and pairwise genetic differentiation. If oestradiols cause widespread feminization of males, then one would expect a reduction in effective population size with concomitant lowering of genetic diversity indices for all loci. Model-based studies of the production of intersex males in roach suggest that in wild populations the presence of intersex males will have a minimal effect upon the population growth rate of roach (Jobling et al. 2002), with some complication brought on by the presence of selective fishing practices (An et al. 2009 in Harris et al. 2011). However, the lack of correlation in effective population size and mean levels of genetic diversity with dissolved levels of oestradiol, do not corroborate the finding of a laboratory study (Harris et al. 2011), using microsatellites, that parenting success is negatively correlated with moderately and severely feminized males. However, Mill Meadow and Stoke-by-Clare were both associated with a recent population bottleneck, which is consistent with the possibility of a significant reduction in the number of effective breeders in the past.

Instead of finding evidence for wholescale reduction of genetic diversity across all loci with EDCs, one observed that a single locus, Lco4, exhibits the greatest correlation of inter-population divergence and difference in oestradiol levels ($p < 0.01$ for all controlled regressions except when the number of tributaries is invoked, albeit near significant ($p = 0.078$)). Such a single-locus pattern may be indicative of selection acting upon a nearby functional locus, presumably exapted to detoxify or excrete the harmful chemical, or to

[40] Calculated using COLONY, a program that employs the sibship method of Wang (Jones and Wang 2009).

increase tolerance to the physiological effects engendered by having oestradiols in the roach's circulatory system. Inspecting the individual locus data further, no single allele of Lco4 showed any significant correlation of frequency with levels of oestradiol, but there was a striking and significant correlation in the frequency of the 197bp allele of the locus Lc290 with oestradiol concentrations, a locus that also displayed a strong predictive relationship with some environmental correlates in the partial Mantel tests.

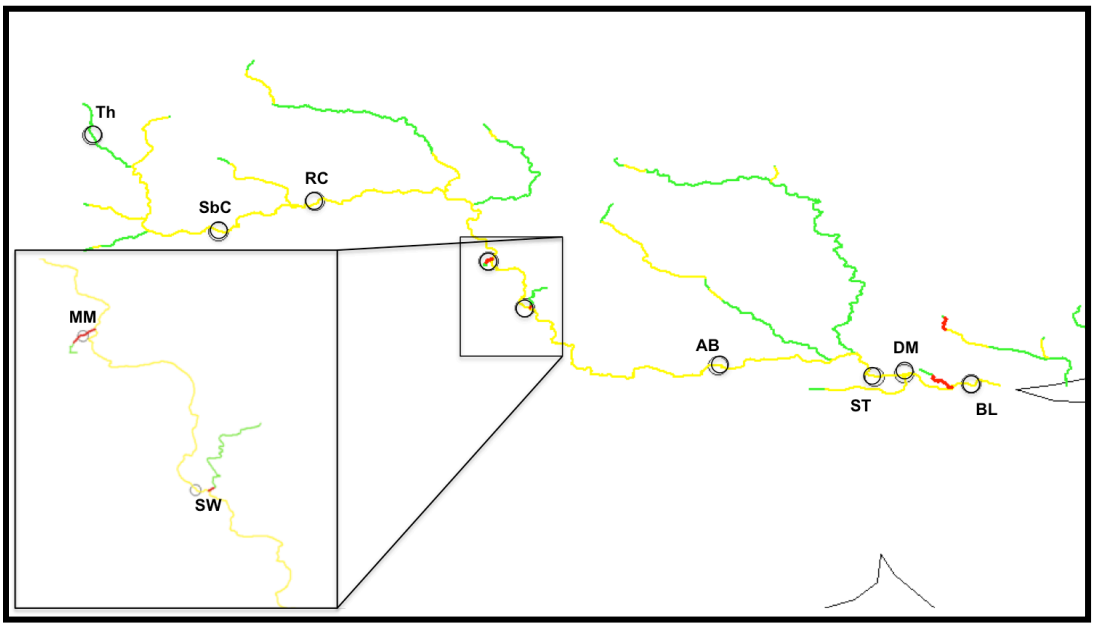


Figure 4.22: Map of the River Stour in which particular stretches are highlighted according to the risk posed by feminizing levels of oestradiols according to the Environment Agency. Green = low; yellow = medium; red = high. Inset: Sample sites of Mill Meadow and Shalford Weir showing their proximity to areas in which oestradiol levels are severe.

It is the only allele to peak significantly with the population at Mill Meadow in the Stour where levels of oestradiols are at least five-fold greater than elsewhere where roach were sampled. Fig 4.23 shows the relationship between oestradiol concentrations and the frequency of the 197bp allele (upper panel) (Pearson's $r = 0.823$, 2-tailed p-value = 0.006). After Shalford Weir there is a consistent relationship between levels of oestradiol and frequency of the allele, with increases and decreases occurring in tandem. What is most noticeable is that there is a spike in frequency of the 197bp allele at the two areas where the concentrations of dissolved oestradiols is greatest (MM: 14.826ng L^{-1}; SbC: 3.227ng L^{-1}, Fig 4.22). That there is little association below Shalford Weir may be because of the mill at Bures, which was deemed extremely hard to bypass, even for the European eel *Anguilla Anguilla* by the Environment Agency. Migration acts against the action of selection, thereby possibly eroding the action of selection downstream of this site.

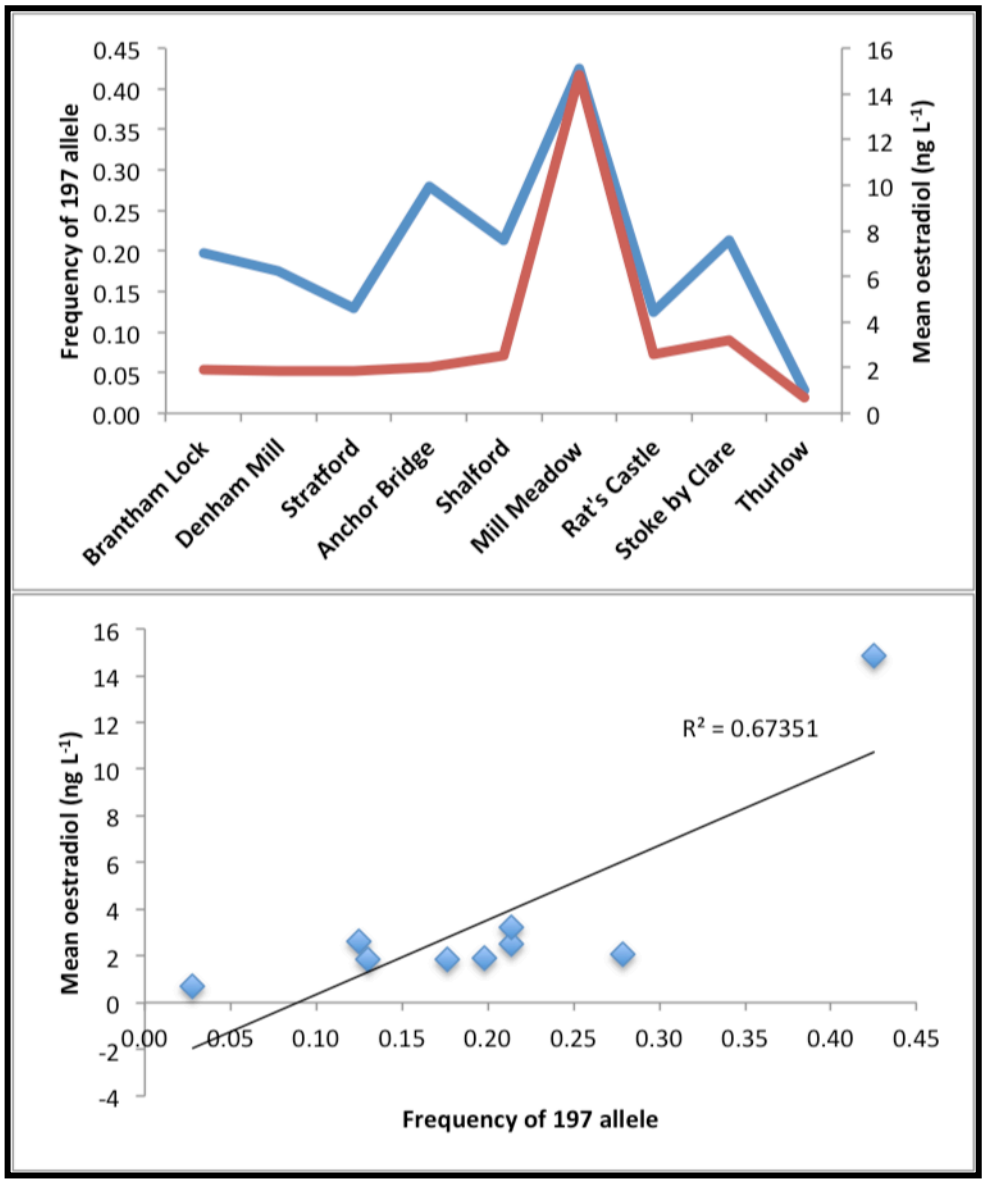


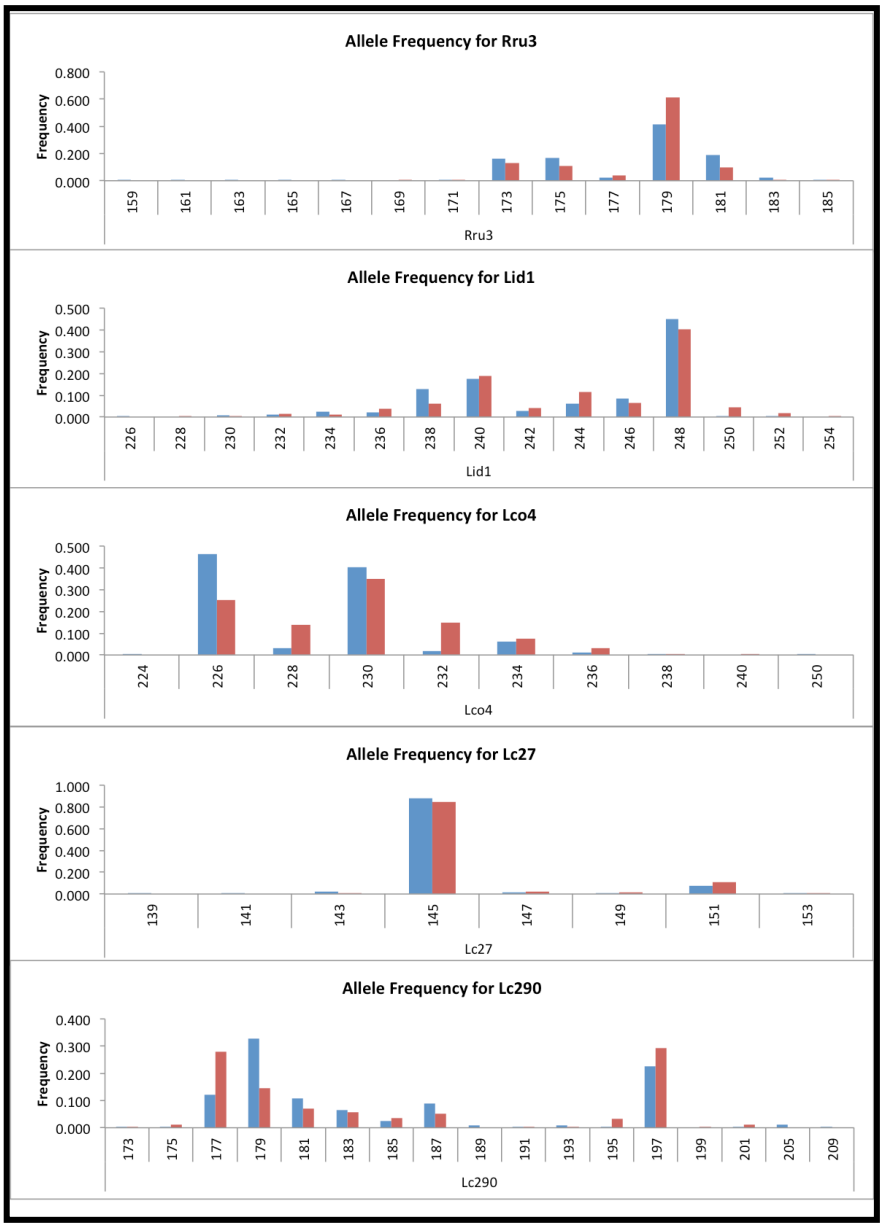
Figure 4.23: Correlation of the 197bp allele (locus Lc290) with mean dissolved oestradiol levels. Top: co-distribution of line graphs showing consonant increases and decreases in both oestradiol levels (red) and allele frequency (blue); bottom: Scatter graph showing the correlation between mean oestradiol levels and allele frequency.

In conclusion, the roach of the Thames and the Stour exhibit similar levels of genetic diversity, but differing patterns and degree of differentiation within their waters. These differences, whilst subtle, will allow a differential approach to managing the genetic stock of roach in these two rivers to be applied to maintain and best preserve the genetic integrity of this important angling commodity.

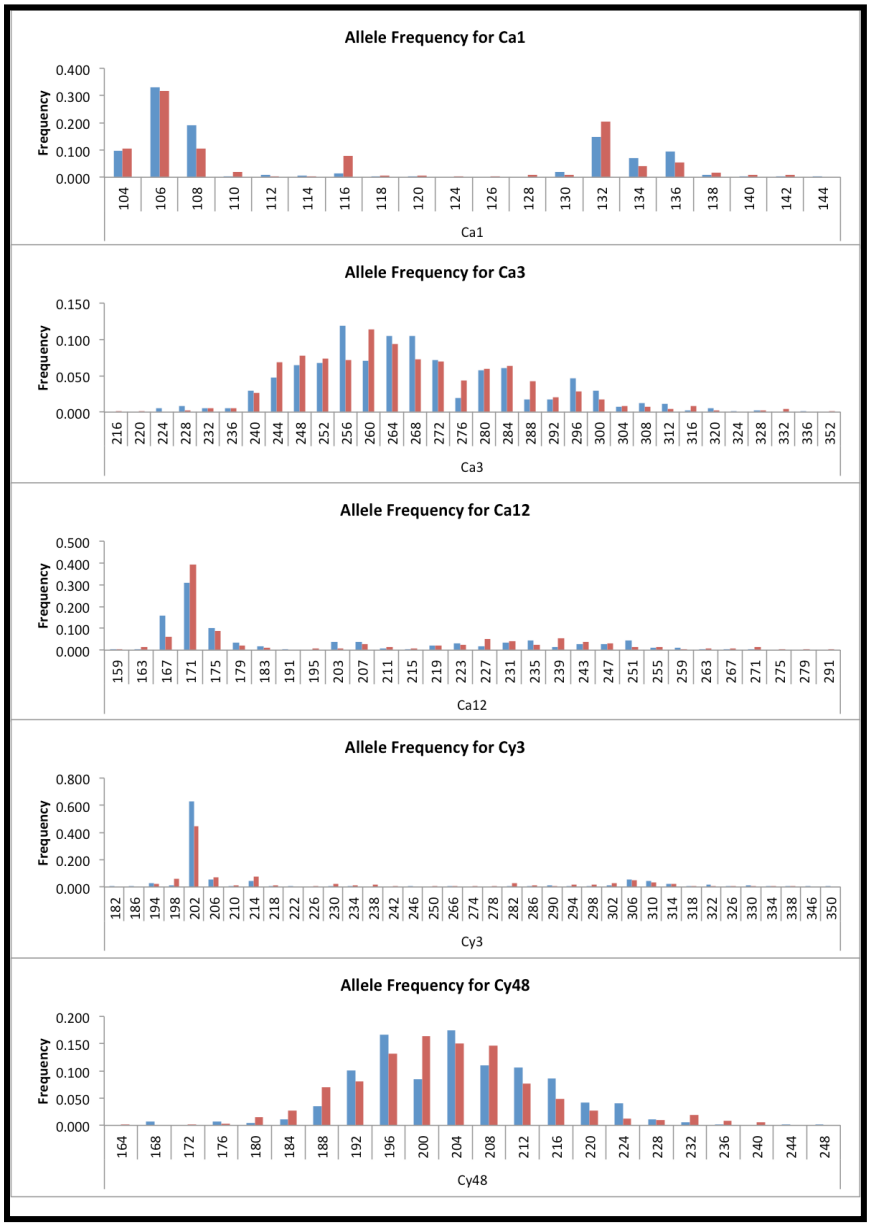
Appendix B

B1 Microsatellite Allele Frequency Data

B1.1: Histograms showing the frequency of each allele found for each locus. Red histograms = Thames; Blue histograms = Stour.



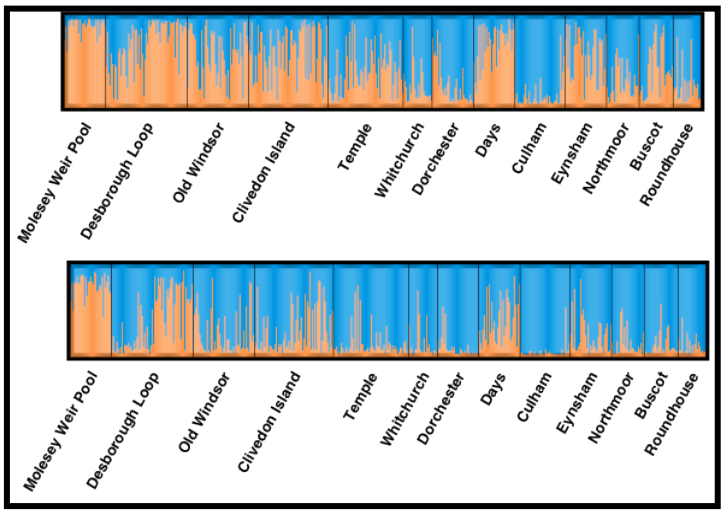
B1.1: Continued.



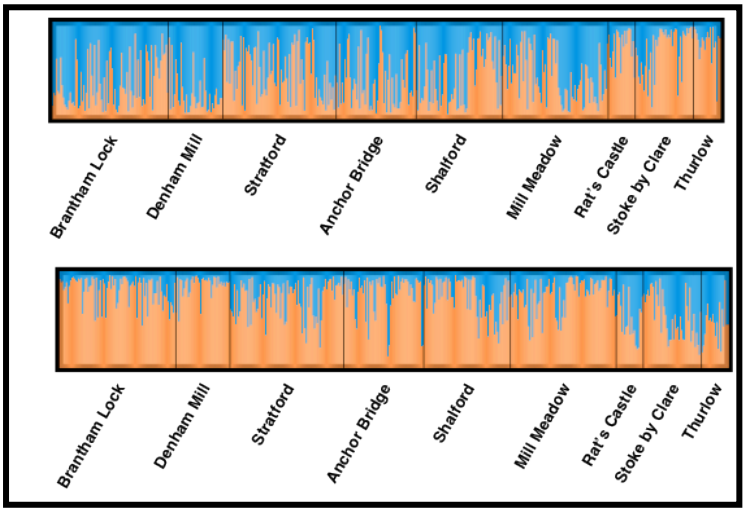
B2 Cryptic Population Structuring

B2.1 Individual STRUCTURE Analyses for the Thames and Stour.

B2.1.1 Results of STRUCTURE analysis of the Thames only – visualized using Distruct 1.1 – with and without prior location information utilized in the analysis (top and bottom panels, respectively). Optimal K = 2.

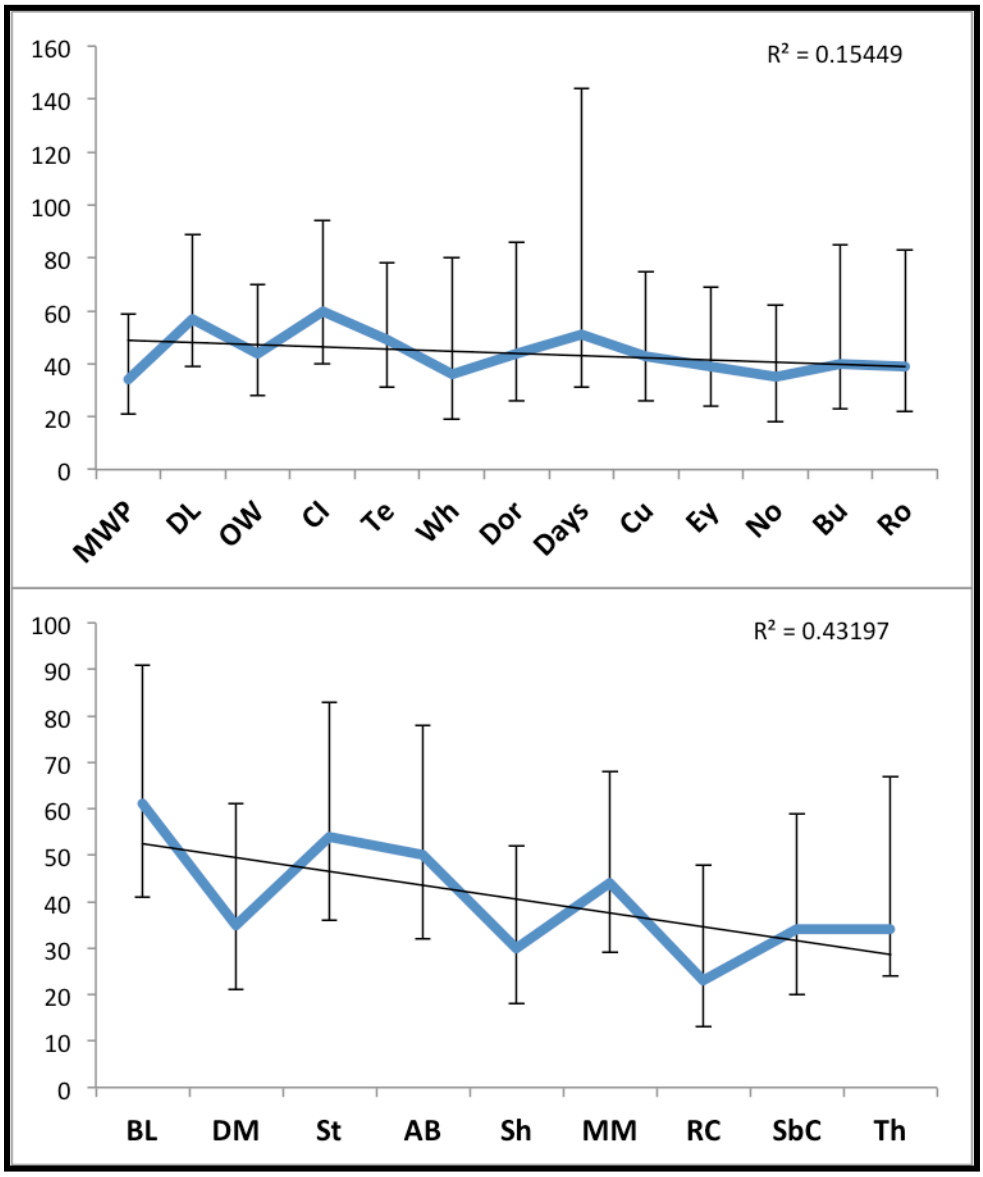


B2.1.2 Results of STRUCTURE analysis of the Stour only – visualized using Distruct 1.1 – with and without prior location information utilized in the analysis (top and bottom panels, respectively). Optimal K = 2.



B3 Effective Population Sizes

B3.1. Line plots of effective population sizes – as determined by the sibship method – for all populations in the Thames (top) and Stour (bottom). 95% CI bars are shown.



Chapter 5: Conclusions - The Management of UK Roach: Inferences to Inform a Coarse Fishery.

"For wild fish in wild places, the [Environment] Agency promotes the principle of self-sustaining fish stocks. These are natural fish populations that develop to suit the aquatic ecosystem they inhabit. Management of such fisheries should be based on knowledge of coarse fish biology. In this way we can ensure that actions are both cost effective and of benefit to all parts of the aquatic ecosystem." **– Anon, Coarse Fish Biology and Management Booklet, Environment Agency.**

The roach in UK waters is characterized by fairly low levels of nucleotide diversity and medium levels of haplotyic diversity, characterizing this species as one that has undergone a geologically recent population bottleneck, followed by population expansion (Grant and Bowen 1998). The overall effect is that quantitative measures of diversity across watersheds reveal limited levels of differentiation as inferred by F_{ST} and G-statistical tests. These results are entirely consistent with European freshwater cypinids found in the northern and western expanses of their ranges. However, some river populations show highly divergent deviations of their D-loop sequences to warrant further attention, for example the populations in the Severn, Sussex Ouse and in the Kennet tributary of the Thames. The latter population is an example of how within-river diversity is greater than that found between rivers, emphasizing the effect that contemporary demographic processes may have on within-river D-loop frequencies in combination with the *in situ* mutation of new haplotypes, albeit in low frequencies, such that inter-river estimates of divergence remain low. On the basis of this study, caution must be applied in applying a purely statistical criterion to combine river populations into single management concerns, as these measures do not take into account within-river diversity, which tended to increase with upstream distance. This result is consistent with that observed with microsatellite loci in linear stretches of the Thames and Stour, whereby the most divergent populations tended to be located upstream (with the exception of MWP in the Thames).

Considering mtDNA further, there is a general increase in mtDNA diversity in roach at lower latitudes in the UK, although the most ancestral haplotype (D4) is found in the north, consistent with mtDNA D-loop data in the grayling (Dawnay et al. 2011). This suggests that although levels of inter-population divergence were on a different scale in the two species, their concordant diversities are suggestive that some freshwater fishes in the UK belong to a cartel of single evolutionary units, with common recent post-glacial histories. However, the implication of this congruence of UK recolonisation history shown by D-loop data from roach and grayling is that the roach ancestral haplotypes found in the Rivers Trent and Tees are probably *not* found there due to anthropogenic translocation, but have arrived naturally via a northerly route of ingress, the route of which is shared by the grayling and probably by other species of obligate freshwater fishes; or that they are remnant of the first wave of colonists that were impacted upon less by subsequent waves of colonizing fish that populations further south and east. The implication is that there may be multiple migratory lineages within the European clade(s) that founded UK freshwater fish diversity, but not enough time has accrued for accumulation of enough neutral sequence variation to make this inference obvious. Additionally, the varying qualitative and quantitative frequencies of rare haplotypes in the smaller rivers and the River Severn suggest that, even before incorporating nDNA data, individual river populations are divergent enough to warrant an ascription of unique repositories of genetic diversity, and therefore should be considered for conservation as single management units.

Most D-loop variation among populations was found to occur within rivers, rather than between them. Similarly, inter-population microsatellite variation was as high within rivers as between rivers, where the pairwise F_{ST} between the pooled populations of the Thames and the Stour was 0.0198, less than the mean F_{ST} among sampled sites within either river. Therefore, like mtDNA diversity, investigators may well misinterpret levels of genetic and ecological differentiation in wild fish populations, if estimates of evolutionary or management potential were based on a naïve surveying of single-point estimates of genetic diversity within a wider habitat, known range or river system. Allied to this point, one cannot over-emphasize the finding that similar genetic outcomes (e.g. overall similar levels of diversity) can be attained by different processes in two closed populations of the same species of freshwater fish. A naïve application of Moritz's allele frequency criterion (1994) would miss the conflicting ecologies and microevolutionary histories of roach in the Thames and the Stour revealed by closer analysis of multiple sites within each river; and, as a result, an ill-informed management policy may be applied. The bottom line of this study is that the mtDNA and the nDNA data suggest that genetically cohesive

populations – or stocks, by Ihssen's definition (1981) – may be found at a hierarchy of spatial scales, from individual headwater populations (e.g. Thurlow in the Suffolk Stour) to entire rivers.

What is clear from the analysis of roach microsatellite DNA is that despite harboring similar overall levels of genetic diversity, the actual distribution of genetic diversity along-river transects of the two southeastern river systems do differ significantly, with more genetic divergence and a higher inferred rate of recent migration within the shorter Suffolk Stour than the longer and more convoluted Thames. The results suggest that within the Stour there are ongoing and localized processes that act against effective migration. In the Stour, the area around Shalford Weir should be investigated further as being of such suitable habitat quality to contribute the majority of migrants for other populations. Because the roach has many spawning qualities in common with other small and mid-sized cyprinids, this area may be important in general as a source of propagules in a classic source-sink metapopulation. Furthermore, there is some evidence for directional selection acting on roach populations at Mill Meadow and Stoke-by-Clare, with similar potential for selection in the highly tidal area of Brantham Lock, both of which need further study. Ecotones within the Rhône drainage were inferred to vary with levels of within-population genetic diversity of allozymes of incumbent roach (Bouvet et al. 1991, 1995), but this is the first known observation of an environmental correlation with microsatellite diversity. Mill Meadow is of particular interest if, for example, locus Lc290 *is* tightly linked to an area in the genome that provides some resistance to the feminizing effects of dissolved oestradiols. Quantitative trait analysis of captive roach in experimental settings using the Lc290 locus as a starting point may help identify other loci that co-vary with an ability to withstand feminization in a pedigree setting. Furthermore, the application of EST-tagged genetic markers to natural populations of roach in areas in which dissolved oestradiols are high is also a desirable focus for further investigation.

The most divergent population in the Stour, at Thurlow, is located in the most lotic part of the system, but crucially also between the inlet and outlet ports for the Ely/Ouse water transport scheme. The potential for extremely low flows in this headwater area may exacerbate population differentiation, as immigration is lowered relative to emigration, possibly accentuating genetic drift. Despite the relatively high numbers of weirs and mills on the Stour, there seems to be little impact upon levels of short-term inter-population migration, although some impact upon levels of genetic differentiation. The old mill at Bures may have impeded migration upstream to some extent, which may disrupt the

ability of the chromosomal haplotype - within which Lc290 resides – to be selected in suitable environments. The presence of IBD, following Kimura & Weiss (1964), was not found in alliance with an increase in heterozygosity or allelic richness in the upstream-downstream direction as predicted, even accounting for the increase in fish numbers with downstream distance (which should translate into increased estimates of heterozygosity (Frankham 1996)). This suggests that some roach actively migrate upstream, acting to counterbalance some loss of diversity, but not enough to combat the action of drift associated with smaller census sizes upstream entirely. In this regard, the roach is different from similar ecological species such as chub (Dehais et al. 2010). This finding is again consistent with a model of a source-sink metapopulation, particularly in the Stour.

The main implication of the above is that roach seem to be less affected by low-impact barriers to migration (e.g. gauging weirs that may be subsumed in flood conditions) than other cyprinids such as chub and even less so than minnows or benthic loaches and sculpins. No high impact barriers (e.g. dams) were studied in this thesis, but the effect upon genetic divergence of allozymes in roach of the Rhône was high (Bouvet et 1991; Laroche et al. 1991), and similarly so for the chub for both allozymes and microsatellites (Laroche et al. 1999; Dehais et al. 2010). However, over the last few years the EA has been constructing a series of fish passes across England & Wales that allow for the upstream migration of eels, salmonids and migratory cyprinids. Fish passes have been found to be particularly effective for faclitiating the upstream migration of roach, especially where pass water velocities are low (Knaepkens et al. 2006). As of 2009, a significant fish pass was constructed at Bures Mill such that migration should no longer be impeded at this location. However, facilitating up-regulated levels of migration to upstream areas of Bures Mill may combat the ability of the upstream populations at Mill Meadow and Stoke-by-Clare to select against the feminizing effects of sewage effluent outflows.

The Thames' roach populations present a different management problem. The Thames has repeatedly, over many centuries, experienced destructive levels of pollution, whereby large areas of the river were made uninhabitable for freshwater fish. The Thames was heavily polluted during the inception of the industrial revolution, especially in the lower reaches below the tidal limit, such that diadromous species were unable to overcome the levels of pollution persistent in its lower course (Mawle & Milner 2003). However, the river is now one of the most ecologically restored systems on the planet (winning an award in 2010 – International Theiss River Award for achievement in river management and

restoration). In 2008, 80% of rivers within the catchment were assayed as either 'good' or 'very good' (as opposed to 53% in 1990) (www.envinronment-agency.gov.uk/news). According to Hughes & Willis (2000), prior to the intense improvement in habitat quality, the Thames was heavily modified, with low habitat diversity, below the confluence of the Thames with the River Kennet (zone "D"), roughly coinciding with a significant genetic break in the present roach dataset. Additionally, the known stocked area of Molesey Weir Pool is genetically divergent, possesses low effective size (consistent with the known retention of low diversity in Swedish roach (Demandt 2010)) and appears to present a genetic break with the nearby upstream population of Desborough Loop. In the year 2000, 30000 fish were released immediately downstream of MWP at two proximate, contiguous sites. If MWP does carry the signature of the released fish, then the retention of an aberrant genetic signal has persisted on a temporal scale similar to that observed in the Swedish study. The ramifications are important, as captively reared fish can show reduced levels of survival and reproduction (Gall 1987). Hughes and Willis (2000) also report that significant water abstraction occurs at Farmoor reservoir, which is located due south from the lock and weir system at Eynsham, a genetically divergent population. If this is significant, the potential for abstraction to correlate with genetically divergent populations of roach is a common occurrence in both rivers.

Other considerations may need to be applied that have not been utilized in the present study, for example morphological variation may be correlated with adaptability to certain ecotones (e.g. Salducci et al. 2004), although previous studies in roach have not found any correlation in genetic divergence and morphology. Moritz (2002) has argued that both historical and adaptive characteristics of natural populations should inform conservation biology. Given the criterion of exchangeability of Crandall et al. (2000), one could not categorically state that roach reciprocally translocated would fare equally well in their new environments. Roach, therefore, on the basis of this principle need to be managed on an individual catchment-by-catchment basis at the very least, consistent with inferences made from the mtDNA evidence. In fact, given the environmental heterogeneity in many catchments, a case could be made for individual rivers within catchments to merit their own protocols, but only mitochondrial evidence can be submitted herein for significant sub-catchment differentiation (e.g. the Thames). In this regard, the genetic data support the potential for independent stocks within rivers, commensurate with Ihssen's concept of a stock with an element of genetic cohension, whilst providing ample evidence for individual watersheds being ringfenced as a very mimum unit for management, under-pinning the fishery stock concept of Smith et al. (1990). However, the nature of the spatial

distribution of equilibrium genetic diversity of roach within both rivers makes delineating 'individual stocks' difficult, but not impossible (e.g. headstream waters are potential repsoitories of divergent, rare and unique diversity – including ecotonic complexity - and warrant protection and further study across a range of coarse fishes).

Where suitable information is available - be it from genetic studies, ecological surveys or by utilizing the working knowledge of anglers and fishing enthusiasts - spawning grounds should be completely protected, because in the absence of costly scientific studies one may not have another means of determining the comparative importance of local spawning grounds to the population within a river as a whole. EU legislation (water and habitat directives) and local parliamentary acts should continue to safeguard - and prosecute against those who damage – high quality littoral riverine habitat. Even the most rudimentary and artificial of habitat improvement schemes can be extremely effective in restoring fish diversity (Wolter 2010). The Fisheries Action Plans (FAPs) of the EA should continue to invest in habitat recuperation and regeneration of extirpated fisheries, but be mindful of the impact of introductions. This study suggests that introduced fish may not integrate into the local roach populous, calling into question the effectiveness of such introductions. Additionally, this study highlights the ability for some areas to replenish neighbouring and distal populations with endogenous propagules. As FAPs are tailored to the needs of individual catchments (Robinson and Whitton 2004), they are in a position to aso implement catchment-specific stocking programs, whereby only autochthonous fishes should be used to replenish local fisheries, if natural replenishment is too slow for the needs of a particular coarse fishery.

As previous studies have pointed out, when it comes to managing fish stocks, through introductions, restorations or translocations, there is no one-size-fits-all policy that can be applied to freshwater fishes (Dehais et al. 2010). Indeed, this study categorically suggests that even within species, population and metapopulation ecologies can differ, which may dramatically alter which management strategy should be applied to be most successful in maintaining healthy, natural and self-sustaining coarse fisheries. The major concluding thesis of this work is that a simple assay of genetic diversity, with either reference to allele or haplotype frequencies, will in many cases overlook subtle microevolutionary processes that may differ across populations within species, and even within small geographical areas in which those species habituate. Moreover, despite high levels of human-mediated translocation, stocking and dispersal, there still exists a significant level of genetic structuring within and/or among all studied river populations of roach. The use of mtDNA

and nDNA markers provided useful insights at different spatial levels that are mutually illuminating; and crucial information would have been lost if either had been excluded from the analysis. The result of this endeavour was an in-depth study of the microevolutionary processes occurring within a species considered by many in the fisheries industry as being far too influenced by human interaction to unveil significant levels of genetic structuring on any scale. The implications are clear: all freshwater fish may be amenable to such studies which may, as in the roach, uncover the idiosyncratic genetic signatures of different environments and ecologies on surprisingly small scales, and upon which action can be taken to protect and conserve freshwater fish biodiversity.

References

Ackroyd P (2009). *The Thames – A Biography*. Doubleday Publishing Ltd. London.

Adams CE (1994). The fish community of Loch Lomond, Scotland; its history and rapidly changing status. *Hydrobiologica*, **290**: 91-102.

Adams CE & Maitland PS (1991). Evidence of further invasions of Loch Lomond by non-native fish species with the discovery of a roach x bream, *Rutilus rutilus* (L.) *x Abramis brama* (L.), hybrid. *Journal of Fish Biology*, **38**(6): 961-963.

Adams CE, Wilson AJ and Ferguson MM (2008). Parallel divergence of sympatric genetic and body size forms of Arctic charr, *Salvelinus alpinus*, from two Scottish lakes. *Biological Journal of the Linnean Society*, **95**: 748-757.

Akaike H (1969). Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics*, **21**: 243-247.

Akaike H (1973). Information theory and an extension of maximum likelihood principle. In: *2nd International Symposium on Information Theory*. Akademia Kiado.

Allendorf FW & Leary RF (1986). Heterozygosity and fitness in natural populations of animals. In: *Conservation Biology: The Science of Scarcity and Diversity* (ed – Soulé ME). Sinauer, Sunderland.

Allendorf FW & Luikart G (2006). *Conservation and the Genetics of Populations*. Blackwell Science Ltd., Oxford.

Almada V & Sousa-Santos C (2010). Comparisons of the genetic structure of *Squalius* populations (Teleostei, Cyprinidae) from rivers with contrasting histories, drainage areas and climatic conditions based on two molecular markers. *Molecular Phylogenetics and Evolution*, **57**: 924-931.

Altukhov PY (1991). The role of balancing selection and overdominance in maintaining allozyme polymorphism. *Genetica*, **85**(1): 79-90.

Alves MJ, Coelho MM and Collares-Pereira MJ (1997). The *Rutilus alburnoides* complex (Cyprinidae): evidence for a hybrid origin. *Journal of Zoological Systematics and Evolutionary Research*, **35**(1): 1-10.

Alves MJ, Coelho H, Collares-Pereira MJ & Coelho MM (2001). Mitochondrial DNA variation in the highly endangered cyprinid fish *Anaecypris hispanica*: importance for conservation. *Heredity*, **87**(4): 463-473.

An W, Hu J, Giesy JP and Yang M (2009). Extinction risk of exploited wild roach (*Rutilus rutilus*) populations due to chemical feminization. *Environmental Science and Technology*, **43**: 7895–7901.

Angers AL & Bernatchez L (1997). Complex evolution of an imperfect salmonid microsatellite and its consequences in inferring allelic differentiation from size information. *Molecular Biology and Evolution*, **14**: 230-238.

Anholt BR (1995). Density dependence resolves the stream drift paradox. *Ecology*, **76**: 2235-2239.

Anon (1577). *The Arte of Angling*. 2nd Facsimile Edition, Princeton University Press, Princeton.

Antao T, Lopes A, Lopes AJ, Beja-Pereira A and Luikart G (2008). LOSITRAN: A workbench to detect molecular adaptation based on a F_{ST}-outlier method. *BMC Informatics*, **9**: 323.

Araújo FG, Bailey RG and Williams PW (1999). Spatial and temporal variations in fish populations in the upper Thames estuary. *Journal of Fish Biology*, **55**: 836-853.

Arthington H and Blühdorn DR (1995). Improved management of exotic aquatic fauna: R&D for Australian rivers. *LWRRDC Occasional Paper, No. 4/95*.

Aston M (1988). Medieval Fish, Fisheries and Fishponds in England. *British Archaeological Reports British Series 182, Oxford*.

Aurelle D & Berrebi P (2001). Genetic structure of brown trout (*Salmo trutta* L.) populations from southwestern France: data from mitochondrial control region variability. *Molecular Ecology*, **10**: 1551-1562.

Avise JC (1995). Mitochondrial DNA polymorphism and a connection between genetics and demography of relevance to conservation. *Conservation Biology*, **9**(3): 686–690.

Avise JC (2000). *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge, MA.

Avise, JC (2004). *Molecular Markers, Natural History and Evolution*, 2nd Ed. Sinauer Associates, Sunderland, MA.

Avise JC, Lansman RA and Shade RO (1979a). The use of restriction endonucleases to measure mitochondrial DNA sequence relatedness in natural populations. I. population structure and evolution in the genus *Peromyscus*. *Genetics*, **92**: 279-295.

Avise JC, Giblin-Davidson C, Larem C, Patton JC and Lansman RA (1979b). Mitochondrial DNA clones and matriarchal phylogeny within and among geographic populations of the pocket gopher, *Geomys pinetis*. *Proceedings of the National Academy of Sciences USA*, **76**: 6694–6698.

Avise JC, Arnold J, and Ball, RM (1987). Intraspecific phylogeography – the mitochondrial-DNA bridge between population genetics and systematics. *Annual Review of Ecological Systematics*, **18**: 489-522.

Avise JC, Ball M and Arnold J (1988). Current versus historical population sizes in vertebrate species with high gene flow: a comparison based on mitochondrial DNA lineages and inbreeding theory for neutral mutations. *Molecular Biology and Evolution*, **5**(4): 331-344.

Ayres DL, Darling A, Zwickl DJ, Beerli P, Holder MT, Lewis PO, Huelsenbeck JP, Ronquist F, Swofford DL, Cummings MP, Rambaut A and Suchard MA (2012). BEAGLE: An application programming interface and high-performance computing library for statistical phylogenetics. *Systems Biology*, **61**(1): 170-173.

Baade U & Fredrich F (1998). Movement and the pattern of activity of the roach in the River Spree, Germany. *Journal of Fish Biology*, **52**: 1165-1174.

Baerwald MR & May B (2004). Characterization of microsatellite loci for five members of the minnow family Cyprinidae found in the Sacramento – San Joaquin Delta and its tributaries. *Molecular Ecology Notes*, **4**: 385-390.

Bahlo M & Griffiths RC (2000). Inference from gene trees in a subdivided population. *Theoretical Population Biology*, **57**: 79-95.

Balding DJ (2003). Likelihood-based inference for genetic correlation coefficients. *Theoretical Population Biology*, **63**: 221-230.

Ballard JW, Melvin RG, Katewa SD and Maas K (2007). Mitochondrial DNA variation is associated with measurable differences in life-history traits and mitochondrial metabolism in *Drosophila simulans*. *Evolution*, **61**(7): 1735-1747.

Balloux F & Lugon-Moulin N (2002). The estimation of population differentiation with microsatellite markers. *Molecular Ecology*, **11**: 155-165.

Bandelt H-J, Forster PG and Röhl (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Ecology*, **16**: 37-48.

Baranyi C, Gollman G and Bobin M (1997). Genetic and morphological variability in roach *Rutilus rutilus*, from Austria. *Hydrobiologia*, **350**(1): 13-23.

Barinova A, Yadrenkina E, Nakajima M and Taniguchi N (2004). Identification and characterization of microsatellite DNA markers developed in ide *Leuciscus idus* and Siberian roach *Rutilus rutilus*. *Molecular Ecology Notes,* **4***:* 86 – 88

Barrett J (2004). Introducing the Murray-Darling Basin native fish strategy and initial steps towards demonstration reaches. *Ecological Management and Restoration*, **5**(1): 15-23.

Bazin E, Glémin S and Galtier N (2006). Population size does not influence mitochondrial genetic diversity in animals. *Science*, **312**: 570-572.

Beaumont MA (1999). Detecting population expansion and decline using microsatellites. *Genetics*, **153**(1): 2013-2029.

Beaumont MA & Nicholls RA (1996). Evaluating loci for use in genetic analysis of population structure. *Proceedings of the Royal Academy of Sciences B*, **263**: 1619-1626.

Beaumont MA & Balding DJ (2004). Identifying adaptive genetic divergence among populations from genome scans, *Molecular Ecology*, **13**: 969-980.

Beaumont MA & Rannala B (2004). The Bayesian revolution in genetics. *Nature Reviews Genetics*, **5**: 251-261.

Beckmann JS and Webber JL (1992). Survey of human and rat microsatellites. *Genomics* **12**: 627-631.

Beebee TJC (2009). A comparison of single-sample effective size estimators using empirical toad (*Bufo calamita*) population data: genetic compensation and population size-genetic diversity correlations. *Molecular Ecology*, **18**: 4790-4797.

Beecham JA & Farnsworth KD (1999). Animal group forces resulting from predator avoidance and competition minimization. *Journal of Theoretical Biology*, **198**(4): 533-548.

Beerli P & Felsenstein J (2001). Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences USA*, **98**: 4563-4568.

Bell MA & Andrews CA (1997). Evolutionary consequences of postglacial colonization of fresh water by primitively anadromous fishes. In: *Evolutionary Ecology of Freshwater Animals* (eds – Stret EB, Städler T and Lively CM). Birkhäuser-Verlag, Basel.

Belle EMS, Piganeau G, Gardner M and Eyre-Walker A (2005). An investigation of the variation in the transition bias among various animal mitochondrial DNA. *Gene*, **355**: 58-66.

Benasson D, De-Xing Z, Hartl DL and Hewitt G (2001). Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology and Evolution*, **16**: 314-321.

Berg MP, Toby Kiers E, Driessen G, van der Heijden M, Kooi BW, Kuenen F, Liefting M, Verhoff HA and Ellers J (2010). Adapt or disperse: understanding species persistence in a changing world. *Global Change Biology*, **16**: 587-598.

Bergek S & Björklund M (2007). Cryptic barriers to dispersal within a lake allow genetic differentiation in the perch. *Evolution*, **61**: 2035-2041.

Bergstrom CT & Pritchard J (1998). Germline bottlenecks and the evolutionary maintenance of mitochondrial genomes. *Genetics*, **149**: 2135-2146.

Bermingham E, Forbes SH, Friedland K and Pla C (1991). Discrimination between Atlantic salmon (*Salmo salar*) of North America and European origin using restriction analyses of mitochondrial DNA. *Canadian Journal of Fisheries and Aquatic Sciences*, **48**: 884-893.

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**: 351-379.

Bernatchez L, Guyomard R and Bonhomme F (1992). DNA sequence variation of the mitochondrial control region among geographically and morphologically remote European brown trout *Salmo trutta* populations. *Molecular Ecology*, **1**: 161-173.

Bernatchez L & Danzmann RG (1993). Congruence in control region sequence and restriction-site variation in mitochondrial DNA of brook charr (*Salvelinus fontinalis* Mitchell). *Molecular Biology and Evolution*, **10**: 1002-1014.

Bernatchez L & Wilson CC (1998). Comparative phylogeography of Nearctic and Palearctic fishes. *Molecular Ecology*, **7**: 431 – 452.

Berners, J (1496). *A Treatyse on fysshnge with an Angle*. Renascence Editions, http://darkwing.uoregon.edu/%7Erbear/berners/berners.html.

Bielawski JP & Gold JR (2002). Mutation patterns of mitochondrial H- and L-strand DNA in closely related cyprinid fishes. *Genetics*, **161**: 1589–1597.

Bijlsma R, Bundgaard J and Boerema AV (2001). Does inbreeding affect the extinction risk of small populations? Predictions from *Drosophila*. *Journal of Evolutionary Biology*, **13**(3): 502-514.

Birky LW (1995). Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proceedings of the National Academy of Sciences USA*, **92**: 11331-11338.

Bjerregaard LB, Korsgaard B and Bjerregaard P (2006). Intersex in wild roach (*Rutilus rutilus*) from Danish sewage effluent-receiving streams. *Ecotoxicology and Environmental Safety*, **64**: 321-328.

Björklund AM, Aho T and Larsson LC (2007). Genetic differentiation in pikeperch (*Sander lucioperca*): the relative importance of gene flow, drift and common history. *Journal of Fish Biology*, **71**: 264-278.

Blanchet S, Rey O, Etienne R, Lek S and Loot G (2010). Species-specific responses to landscape fragmentation: implications for management strategies. *Evolutionary Applications*, **3**(3): 291-304.

Blasco-Costa I, Waters JM and Pulin R (2011). Swimming against the current: genetic structure, host mobility and the drift paradox in trematode parasites. *Molecular Ecology*, **21**(1): 207-217.

Blum MJ, Bagley MJ, Walters DM, Jackson SA, Bernard Daniel F, Chaloud DJ & Cade BS (2011). Genetic diversity and diversity of stream fishes covary across a land-use gradient. *Oecologia*, DOI: 10.1007/s00442-011-2078-x.

Bohlen J, Slechtova V, Doadrio I and Rab P (2007). Low mitochondrial divergence indicates a rapid expansion across Europe in the weather loach, *Misgurnus fossilis* (L.). *Journal of Fish Biology*, **71**: 186-194.

Bŏhme M & Ilg A (2003). Fossil Fishes, Amphibians, Reptiles – fosFARbase. (http://ww.wahre-staerke.com).

Bonnet E & Van de Peer Y (2002). Zt: a software tool for simple and partial Mantel tests. *Journal of Statistical Software*, **7**: 1-12.

Bossart JL & Powell DP (1998). Genetic estimates of population structure and gene flow: limitations, lessons and new directions. *Trends in Ecology and Evolution*, **13**: 202-206.

Bouvet Y Pattee E and Meggouh F (1985). The contribution of backwaters to the ecology of fish populations in a large river. Preliminary results on fish migrations within a side arm and from the side arm to the main channel of the Rhone. *Verh. int. Ver. Limnol.*, **22**: 2576-2580.

Bouvet Y, Soewardi K and Pattee E (1991). The discrimination of roach *Rutilus rutilus* (Linnaeus, 1758) populations in different parts of a river system. An investigation using biochemical markers. *Hydrobiologia*, **209**: 161-167.

Bouvet Y, Bobin M, Maslin JL and Pattee (1995). The genetic structure of roach populations in two contrasted large rivers. *Hydrobiologia*, **303**: 229-234.

Brassington RA & Ferguson A (1976). Electrophoretic identification of roach (*Rutilus rutilus* L.), rudd (*Scardinius erythropthalmus* L.), bream (*Abramis brama* L.) and their natural hybrids. *Journal of Fish Biology*, **9**(6): 471-477.

Broquet T & Petit E (2009). Molecular estimation of dispersal for ecology and population genetics. *Annual Review of Ecology, Evolution and Systematics*, **40**: 193-216.

Broughton RE, Milam JE and Roe BA (2001). The complete sequence of the zebrafish (*Danio rerio*) mitochondrial genome and evolutionary patterns in vertebrate mitochondrial DNA. *Genome Research*, **11**(11): 1958-1967.

Brown BL & Swan CM (2010). Dendritic network structure constrains metacommunity properties in riverine ecosystems. *Journal of Animal Ecology*, **79**: 571-580.

Brown GG, Gadaleta G, Pepe G, Saccone C and Sbisa E (1986). Structure conservation and variation in the D-loop containing region of vertebrate mitochondrial DNA. *Journal of Molecular Biology*, **192**: 503-511.

Brown W (1985). The mitochondrial genome of animals. In: *Molecular Evolutionary Genetics* (ed - Macintyre RJ). Plenum, New York.

Bryja J, Smith C, Konečný A and Reichard M (2010). Range-wide population genetic structure of the European bitterling (*Rhodeus amarus*) based on microsatellite and mitochondrial DNA analysis. *Molecular Ecology*, **19**: 4708-4722.

Callen D, Thompson A and Shen Y (1993). Incidence and origin of 'null' alleles in the (AC) microsatellite markers. *American Journal of Human Genetics*, **52**: 922-927.

Campbell RN (1971). The growth of brown trout *Salmo trutta* L. in northern Scottish lochs with special reference to the improvement of fisheries. *Journal of Fish Biology*, **3**: 1-28.

Cantatore P, Roberti M, Pesole A, Ludovico F, Milella M, Gadaleta N and Saccone C (1994). Evolutionary analysis of cytochrome b sequences in some perciformes: evidence for a slower rate of evolution in mammals. *Journal of Molecular Evolution*, **39**: 589-597.

Carlsson J (2008). Effects of microsatellite null alleles on assignment testing. *Journal of Heredity*, **99**(6): 616-623.

Carr SM & Marshal DH (2008). Intraspecific phylogeographic genomics from multiple complete mitochondrial genomes in Atlantic cod (*Gadus morhua*): origins of the "codmother," transatlantic vicariance and midglacial population expansion. *Genetics*, **180**: 381-389.

Carrott J, Hall A, Issitt M, Kenward H and Large F (1994). Assessment of biological remains from excavations at Withow Gap, Skipsea, Humberside (site code: CAS489). *Reports from the Environmental Archaeology Unit, York*, **94/37**.

Carter M (2004). In: *Freshwater Fishes in Britain* (eds – Davies, C, Shelley J, Harding P, McLean I, Gardner R and Peirson G). Harley Books Ltd., Colchester.

Carvalho GR (1993). Evolutionary aspects of fish distribution: genetic variability and adaptation. *Journal of Fish Biology*, **43** (Supplement A): 53-73.

Carvalho GR & Hauser L (1994). Molecular genetics and the stock concept in fisheries. *Reviews in Fish Biology and Fisheries* **4**: 326-350.

Cavalli-Sforza LL & Edwards AWF (1967). Phylogenetic analysis: methods and estimation procedures. *American Journal of Human Genetics*, **19**: 233-257.

Cavender TM (1991). The fossil record of Cyprinidae. In: *Cyprinid Fishes: Systematics, Biology and Exploitation* (eds – Winfield IJ and Nelson JE). Chapman & Hall, London.

Chakraborty R, de Andrade M, Daiger S and Budowle B (1992). Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications. *Annals of Human Genetics*, **56**: 45-57.

Champion AS & Swain S (1974). A note on the movements of coarse fish passing through the Ministry's trapping installation on the River Axe, Devon. *Journal of the Institute of Fisheries Management*, **5**: 89-92.

Chappell J (1974). Relationships between sea levels, ^{18}O variations and orbital perturbations during the past 250000 years. *Nature*, **252**: 199-202.

Chapuis M-P & Estoup A (2007). Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution*, **24**(3): 621-631.

Charlesworth B (1998). Measures of divergence between populations and the effect of forces that reduce variability. *Molecular and Biological Evolution*, **15**: 538-543.

Chaturvedi A, Mohindra V, Singh RK, Lal KK, Punia P, Bhaskar R, Mandal A, Narain L and Lakra WS (2011). Population genetic structure and phylogeography of cyprinid fish, *Labeo dero* (Hamilton 1822) inferred from allozyme and microsatellite DNA marker analysis. *Molecular Biology Reports*, **38**: 3513-3529.

Chikhi L, Sousa VC, Luisis P, Goossens B and Beaumont MA (2010). The confounding effects of population structure, genetic diversity and the sampling scheme on the detection and quantification of population size changes. *Genetics*, **186**: 983-995.

Chistiakov DA, Hellemans B and Volckaert AM (2006). Microsatellites and their genomic distribution, evolution, function and applications: a review with special reference to fish genetics. *Aquaculture*, **255**: 1-29.

Ciofi C, Beaumont MA, Swingland IR and Bruford MW (1999). Genetic divergence and units for conservation in the Komodo Dragon *Varanus komodiensis*. *Proceedings of the Royal Society B*, **266**(1435): 2269-2274.

Clement M, Posada D and Crandall KA (2000). TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**: 1657-1659.

Cochran RG (1954). Some methods for strengthening the common χ^2 tests. *Biometrics*, **10**: 417-451.

Coelho MM & Zalewski M (1995). Evolutionary adaptations by fish to ecotonal complexity in spatially variable landscapes – a perspective. *Hydrobiologia*, **303**: 223-228.

Conner JK & Hartl DL (2004). *A Primer of Ecological Genetics*. Sinauer, Sunderland, Massachusetts, USA.

Cooke SJ & Cowx IG (2006). Contrasting recreational and commercial fishing: searching for common issues to promote unified conservation of fisheries resources and aquatic environments. *Biological Conservation*, **128**: 93-108.

Copp GH (1990). Shifts in the microhabitat of larval and juvenile roach *Rutilus rutilus* (L.) in a floodplain channel. *Journal of Fish Biology*, **36**: 683-692.

Copp GH (1992). Comparative microhabitat use of cyprinid larvae and juveniles in a lotic floodplain channel. *Environmental Biology of Fishes*, **33**: 181-193.

Copp GH (1997). Importance of marinas and off-channel water bodies as refuges for young fishes in a regulated river. *Regulated Rivers: Research and Management*, **13**:303-307.

Copp GH & Jurajda P (1993). Do small riverine fish move inshore at night? *Journal of Fish Biology*, **43** (supplement A): 229-241.

Corander J Waldmann P, Marttinen P and Sillanpää (2004). BAPS 2: enhanced possibilities for the analysis of genetic population structure. *Bioinformatics*, **20**(15): 2363-2369.

Cornuet J, Piry S, Luikart G, Estoup A and Solignac M (1999). New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics*, **153**: 1989-2000.

Costedoat C, Chappaz R, Barascud B Guillard G and Gilles A (2006). Heterogeneous colonization pattern of European Cyprinids, as highlighted by the dace complex (Teleostei: Cyprinidae). *Molecular Phylogenetics and Evolution*, **41**: 127-148.

Costello AB, Down TE, Pollard SM, Pacas CJ and Taylor EB (2003). The influence of history and contemporary stream hydrology on the evolution of genetic diversity within species: an examination of microsatellite DNA variation in bull trout, *Salvelinus confluentus* (Pisces: Salmonidae). *Evolution*, **57**(2): 328-344.

Cotgreave P (1993). The relationship between body-size and population abundance in animals. *Trends in Ecology & Evolution*, **8**: 244-248.

Cowx IG (1983). The biology of bream, *Abramis brama* (L.), and its natural hybrid with roach, *Rutilus rutilus* (L.), in the River Exe. *Journal of Fish Biology*, **22**: 631-646.

Cowx IG (1998). *Stocking and Introduction of Fish*. Fishing News Books, Blackwell Science Ltd., Oxford.

Cowx IG (2002). Recreational fishing. In: *Handbook of Fish Biology and Fisheries*, Vol. 2. (eds – Hart PJB and Reynolds JD). Blackwell Science Ltd., Oxford.

Cowx IG and Gerdeaux D (2004). The effects of fisheries management practices on freshwater ecosystems. *Fisheries Management and Ecology*, **11**: 145-151.

Crandall KA & Templeton AR (1993). Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics*, **134**: 959-969.

Crandall KA, Bininda-Emonds RP, Mace GM and Wayne RK (2000). Considering evolutionary processes in conservation biology. *Trends in Ecology and Evolution*, **15**(7): 290-295.

Cristescu R, Sherwin WB, Handasyde K, Cahill V and Cooper DW (2010). Detecting bottlenecks using BOTTLENECK 1.2.02 in wild populations: the importance of microsatellite structure. *Conservation Genetics*, **11**: 1043-1049.

Crow JF & Kimura M (1970). An *Introduction to Population Genetics Theory*. Burgess Publishing Company, Minneapolis, MN.

Crow JF & Aoki K (1984). Group selection for a polygenic behavioural trait: estimating the degree of population subdivision. *Proceedings of the National Academy of Sciences USA*, **81**(19): 6073-6077.

Culling MA, Janko K, Boron A, Vasil'Ev VP, Cote IM and Hewitt GM (2006). European colonisation by the spined loach (*Cobitis taenia*) from Ponto - Caspian refugia based on mitochondrial DNA variation. *Molecular Ecology*, **15**: 173-190.

Dakin EE & Avise JC (2004). Microsatellite null alleles in parentage analysis. *Heredity*, **93**: 504-509.

Davies N, Villablanca FX and Roderick GK (1999). Determining the source of newly founded populations: multilocus genotyping in nonequilibrium population genetics. *Trends in Ecology and Evolution*, **14**: 17-21.

Dawnay N, Dawnay L, Hughes RN, Cove R and Taylor MI (2011). Substantial genetic structure among stocked and native populations of the European grayling (*Thymallus thymallus*, Salmonidae) in the United Kingdom. *Conservation Genetics*, **12**: 731-744.

De La Peña A (1995). Tertiary fishes from the Iberian Continental Basins: History and fossil record. *Coloquios de Paleontología*, **47**: 25-47.

De Ligny W (1969). Serological and biochemical studies in fish populations. *Oceanography and Marine Biology – An Annual Review,* **7**: 411-513.

DEFRA (2010). *Eel Management Plans for the United Kingdom: Anglian River Basin District*. Department for Environment, Food and Rural Affairs, UK Government.

Dehais C, Eudeline R, Berrebi P and Argillier C (2010). Microgeographic genetic isolation in chub (Cyprinidae: *Squalius cephalus*) population of the Durance River: estimating fragmentation by dams. *Ecology of Freshwater Fish*, **19**: 267-278.

Demandt M (2010). Temporal changes in genetic diversity of isolated populations of perch and roach. *Conservation Genetics*, **11**(1): 249-255.

Demandt M & Björkland M (2007). Loss of genetic variability in reintroduced roach (*Rutilus rutilus*) populations. *Journal of Fish Biology*, **70** (Supplement B): 225 – 261.

Dempster AP, Laird NM and Rubin DB (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Statistical Society Series B*, **39**(1): 1-38.

DeWoody JA & Avise JC (2000). Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. *Journal of Fish Biology*, **56**(3): 461-473.

DeYoung RW & Honeycutt RL (2005). The molecular toolbox: genetic techniques in wildlife ecology and management. *Journal of Wildlife Management*, **69**: 1362-1384.

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M and Freimer NB (1994) Mutational processes of simple-sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences USA*, **91**: 3166–3170.

Diamond M (1985). Some observations on spawning by roach *Rutilus rutilus* L., and bream *Abramis brama* L., and their implications for management. *Aquaculture Research*. **16**(4): 359-367.

Dimmick WW, Berendzen PB and Golubtsov AS (2001). Genetic comparison of three *Barbus* (Cyprinidae) morphotypes from the Genale River Ethiopia. *Copeia*, **2001**(4): 1123-1129.

Dimsoski P, Toth G and Bagley M (2000). Microsatellite characterisation in central stoneroller *Campostoma anomalum* (Pisces: Cyprinidae). *Molecular Ecology*, **9**: 2187-2189.

Dobzhansky T (1962). *Oral History Memoir*. Columbia University, Oral History Research Office, New York.

Douglas MR, Brunner PC and Douglas ME (2003). Drought in an evolutionary context: molecular variability in Flannelmouth Sucker (*Catostomus latipinnis*) from the Colorado River of western North America. *Freshwater Biology*, **48**(7): 1254-1273.

Dowling TE, Minckley WL and Marsh PC (1996). Mitochondrial DNA diversity within and among populations of razorback sucker (*Xyrauchen texanus*) as determined by restriction endonuclease analyses, *Copeia*, **3**: 542-550.

Drummond A, Ho S, Phillips M and Rambaut A (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biology*, **4**: 88.

Drummond AJ, Suchard MA, Xie D and Rambaut A (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**: 1969-1973.

Dubut V, Fouquet A, Voisin A, Costedoat C, Chappaz R and Gilles R (2012). From late Miocene to Holocene: processes of differentiation within the *Telestes* genus (Actinopterygii: Cyprinidae). *PLoS ONE*: **7**(3): e34423.

Durand JD, Persat H & Bouvet Y (1999). Phylogeography and postglacial dispersion of the chub (*Leuciscus cephalus*) in Europe. *Molecular Ecology*, **8**: 989-997.

Dyer C (1988). The consumption of freshwater fish in Medieval England. In: *Medieval Fish, Fisheries and Fishponds in England* (ed – Aston M). British Archaeological Reports British Series 182, Oxford.

Earl DA & van Holdt BM (2011). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, DOI: 10.1007/s12686-011-9548-7

Edwards YJ, Elgar G, Clark MS, and Bishop MJ (1998). The identification and characterisation of microsatellite s in the compact genome of the Japanese pufferfish, *Fugu rupripes*: perspectives in functional and comparative genomic analysis. *Journal of Molecular Biology*, **278**: 843-854.

Ehinger M, Fontinillas P, Petit E and Perrin N (2002). Mitochondrial DNA variation along an altitudinal gradient in the greater white-toothed shrew, *Crocidura russula*. *Molecular Ecology*, **11**: 939-945.

Ellegren H (2000a). Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in Genetics*, **16**: 551-558.

Ellegren H (2000b). Heterogeneous mutation processes in human microsatellite DNA sequences. *Nature Genetics*, **24**: 400-402.

Ellegren H (2004). Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics*, **5**: 435-445.

Elmer KR, van Houdt JKJ, Meyer A and Volckaert FAM (2008). Population genetic structure of North American burbot (*Lota lota maculosa*) across the Nearctic and its contact zone with Eurasian burbot (*Lota lota lota*). *Canadian Journal of Fisheries and Aquaculture*, **65**: 2412-2426.

Engelbrecht CC, Freyhof J, Nolte A, Rassmann K, Schliewen U and Tautz D (2000). Phylogeography of the bullhead *Cottus gobio* (Pisces: Teleostei: Cottidae) suggest a pre-Pleistocene origin of the major central European populations. *Molecular Ecology*, **9**: 709-722.

Environment Agency (1999). *Coarse Fisheries Strategy*. Bristol, UK: Environment Agency.

Environment Agency (2010). *River Basin Management Plan Anglian River Basin District*. Peterborough, UK: Environment Agency.

Eschelle AF, van Den Bussche RA, Malloy Jr. TP, Haynie ML and Minckley CO (2000). Mitochondrial DNA variation in pupfishes assigned to the species *Cyprinodon macularius* (Atherinomorpha: Cyprinodontidae): taxonomic implications and conservation genetics. *Copeia*, **2000**(2): 353-364.

Estoup A and Angers B (1998). Microsatellites and minisatellites for molecular ecology, theoretical and empirical considerations. In: *Advances in Molecular Ecology* (ed – Carvalho GR). IOS Press, Amsterdam.

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**(8): 2611-2620.

Evans SG, Spillett PB and Colquhoun K (2003). Southeast housing development – the quest for sustainability: water and sewerage needs. *Water and Environment Journal*, **17**: 257-261.

Everard M (2006). *The Complete Book of the Roach*. Medlar Press Ltd., Ellesmere.

Evlanov IA (1986). Study of the intrapopulational structure of the roach *Rutilus rutilus* in Lake Vishtynetskoe, Russian SFSR. *Vopr. Ikhtiol*. **26**: 398-401.

Ewens WJ (1982). On the concept of effective population size. *Theoretical Population Biology*, **3**: 87-112.

Excoffier L, Smouse PE and Quattro JM (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**: 479-491.

Excoffier L, Laval G and Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**:47-50.

Excoffier L & Heckel G (2006). Computer programs for population genetics data analysis: a survival guide. *Nature Reviews*, **7**: 745-758.

Eyre-Walker A (2000). Do mitochondria recombine in humans? *Philosophical Transactions of the Royal Society B*, **355**: 1573-1580.

Fagan WF (2002). Connectivity, fragmentation and extinction risk in dendritic metapopulations. *Ecology*, **83**: 3243–3249.

Fagan WF & Holmes EE (2006). Quantifying the extinction vortex. *Ecology Letters*, **9**: 51-60.

Fagan WF, Grant EHC, Lynch HJ and Unmack PJ (2009). Riverine landscapes: ecology for an alternative geometry. In: *Spatial Ecology* (eds – Ruan S, Cosner C and Cantrell S. Chapman and Hall, Berlin.

Falush D, Stephens M and Pritchard J (2003). Inference of population structure: Extensions to linked loci and correlated allele frequencies. *Genetics*, **164**: 1567-1587.

Faubet P & Gaggiotti OE (2008). A new Bayesian method to identify the environmental factors that influence recent migration. *Genetics*, **178**: 1491–1504.

Felsenstein J (1989). PHYLIP Phylogenetic Inference Package (Version 3.2). *Cladistics*, **5**: 164-166.

Felsenstein J (1993). PHYLIP (Phylogenetic Inference Package) user manual. Department of Genetics, University of Washington, Seattle.

Fisher RA (1930). *The Genetical Theory of Natural Selection*. Oxford University Press, Oxford.

Fitch WM (1967). Evidence suggesting a non-random character to nucleotide replacements in naturally occurring mutations. *Journal of Molecular Biology*, **26**: 499-507.

Fitzmaurice P (1974). Occurrence of roach, *Rutilus rutilus* L. in the Shannon catchment. *Irish Naturalist's Journal*, **18**: 19.

Fitzmaurice P (1984). The effects of freshwater fish introductions into Ireland. FAO Report.

Foote AD, Morin PA, Durban JW, Pitman RL, Wade P, Willerslev E, Thomas M, Gilbert P and de Fonseca RR (2011). Positive selection on the killer whale mitogenome. *Biology Letters*, **7**(1): 116-118.

Foran DR, Hixson JE and Brown WM (1988). Comparisons of ape and human sequences that regulate mitochondrial DNA transcription and D-loop synthesis. *Nucleic Acids Research*, **16**: 5841-5861.

Frankham R (1995). Effective population size/census population size ratios in wildlife: a review. *Genetical Research*, **66**: 85-107.

Frankham R (1996). Relationship of genetic variation to population size in wildlife. *Conservation Biology*, **10**: 1500-1508.

Frankham R, Ballou JD and Briscoe DA (2003). *An Introduction to Conservation Genetics*. Cambridge University Press, Cambridge.

Franklin IA (1980) Evolutionary changes in small populations. In: *Conservation Biology: An Evolutionary and Ecological Perspective* (eds - Soulé M & Wilcox BA) Sinauer Associates, Sunderland, Massachusett, USA.

Fraser DF, Lippé C and Bernatchez L (2004). Consequences of unequal population size, asymmetric gene flow and sex-biased dispersal on population structure in brook charr (*Salvelinus fontinalis*). *Molecular Ecology*, **13**: 67-80.

Fu Y-X (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**: 915-925.

Fu Y-X & Li W-H (1993). Statistical tests of neutrality of mutations. *Genetics*, **133**: 693-709.

Gaggiotti O & Excoffier L (2000). A simple method of removing the effect of a bottleneck and unequal population sizes on pairwise genetic differences. *Proceedings of the Royal Society London B*, **267**: 81-87.

Gall GAE (1987). Inbreeding. In: *Population Genetics and Fisheries Management* (eds – Ryman N and Utter F). University of Washington Press, Seattle, USA.

Galtier N, Enard D, Radondy Y, Bazin E and Belkhir K (2006). Mutation hot spots in mammalian mitochondrial DNA. *Genome Research*, **16**: 215-222.

Galtier N, Nabholz B, Glemin S and Hurst GDD (2009). Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology*, **18**: 4541-4550.

Garner P (1996). Diel behaviour of juvenile 0-group fishes in a regulated river: the Great Ouse, England. *Ecology of Freshwater Fish*, **5**: 175-182.

Gauldie RW (1988). Tagging and genetically isolated stocks of fish: a test of one stock hypothesis and the development of another. *Journal of Applied Ichthyology*, **4**: 168-173.

Genty D, Blamart D, Ouahdi R, Gilmour M, Baker A, Jouzel J and Van-Exter S (2003). Precise dating of Dansgaard-Oeschger climate oscillations in western Europe from stalagmite data. *Nature*, **241**: 833-837.

George D & Blieck A (2011) Rise of the Earliest Tetrapods: An Early Devonian Origin from Marine Environment. *PLoS ONE* **6**(7): e22136. doi:10.1371/journal.pone.0022136

Gering EJ, Opazo JC and Storz JF (2009). Molecular evolution of cytochrome b in high- and low-altitude deer mice (genus *Peromyscus*). *Heredity*, **102**: 226-235.

Gilles A, Lecointre G, Faure E, Chappaz R and Brun G (1998). Mitochondrial phylogeny of the European cyprinids: implications for their systematic, reticulate evolution, and colonization time. *Molecular Phylogenetics and Evolution*, **10**(1): 132-143.

Gilles A, Lecointre G, Miquelis A, Loerstcher M, Chappaz R and Brun G (2001). Partial combination applied to phylogeny of the European cyprinids using the mitochondrial control region. Molecular *Phylogenetics and Evolution*, **19**(1): 22-33.

Gillespie JH (1974). Polymorphism in patchy environments. *American Naturalist*, **108**: 145-151.

Gillet C & Dubois JP (1995). A survey of the spawning of perch (*Perca fluviatilis*), pike (*Esox lucius*), and roach (*Rutilus rutilus*), using artificial spawning substrates in lakes. *Hydrobiologia*, **300/301**: 409-415.

Gillet C & Quétin P (2006). Effect of temperature changes on the reproductive cycle of roach in Lake Geneva from 1983 to 2001. *Journal of Fish Biology*, **69**: 518-534.

Gilpin ME & Soulé ME (1986). Minimum viable populations: processes of species extinction. In: *Conservation Biology: The Science of Scarcity and Diversity* (ed – Soulé ME). Sinauer, Sunderland, USA.

Gojobori T, Li W-H and Graur D (1982). Patterns of nucleotide substitution in pseudogenes and functional genes. *Journal of Molecular Evolution*, **18**(5): 360-369.

Goldberg EE, Lynch HJ, Neubert MG and Fagan WF (2010). Effects of branching spatial structure and life history on the asymptotic growth rate of a population. *Theoretical Ecology*, **3**: 137-152.

Goldspink CR (1977). The return of marked roach (*Rutilus rutilus* L.) to spawning grounds in Tjeukemeer, The Netherlands. *Journal of Fish Biology*, **11**: 599-603.

Gollman G, Bouvet Y, Karakoussis Y and Triantaphyllidis C (1997). Genetic variability in *Chondrostoma* from Austrian, French and Greek rivers (Teleostei, Cyprinidae). *Journal of Zoological Systematics and Evolutionary Research*, **35**(4): 165-169.

Goudet J (1995). FSTAT (Version 1.2): A Computer program to calculate *F*-Statistics. *Journal of Heredity*, **86**(6): 485-486

Goudet J, Raymond M, de Meeüs T and Rousset F (1996). Testing differentiation in diploid populations. *Genetics*, **144**: 1931-1938.

Grant WS & Bowen BW (1998). Shallow population histories in deep evolutionary lineages of marine fishes: insights from sardines and anchovies and lessons for conservation. *Journal of Heredity*, **89**(5): 415-426.

Graur D & Li W-H (2000). *Fundamentals of Molecular Evolution*. Sinauer, Sunderland, MA.

Griekspoor A & T Groothius. 4Peaks version 1.7. www.mekentosj.com.

Griffiths RC & Li W-H (1983). Simulating allele frequencies in a population and the genetic differentiation of populations under mutation pressure. *Theoretical Population Biology*, **23**: 19-33.

Guillot G (2008). Inference of structure in subdivided populations at low levels of genetic differentiation. The correlated allele frequencies model revisited. *Bioinformatics*, **24**(19): 2222-2228.

Guillot G, Mortier F and Estoup A (2005). Geneland: A program for landscape genetics. *Molecular Ecology Notes*, **5**: 712-715.

Guindon S & Gascuel O (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, **52**: 696-704.

Gum B, Gross R and Kuehn R (2005). Mitochondrial and nuclear DNA phylogeography of European grayling (*Thymallus thymallus*): evidence for secondary contact zones in central Europe. *Molecular Ecology*, **14**: 1707-1725.

Gum B, Gross R and Geist J (2009). Conservation genetics and management implications for European grayling, *Thymallus thymallus*: synthesis of phylogeography and population genetics. *Fisheries Management and Ecology*, **16**(1): 37–51.

Guo SW & Thompson EA (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*, **48**(2): 361-372.

Guo X & Guo Q (2006). Mutations in the mitochondrial DNA D-loop region occur frequently in human osteosarcoma. *Cancer Letters*, **239**(1): 151-155.

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

Habib M, Lakra WS, Mohindra V, Khare P, Barman AS, Singh K, Lal KK, Punia P and Khan AA (2011). Evaluation of cytochrome b mtDNA sequences in genetic diversity studies of *Channa marulius* (Channidae: Perciformes). *Molecular Biology Reports*, **38**: 841-846.

Hale PA (1958). Roach, *Rutilus rutilus*, L. occurring in Co. Tyrone. *Irish Naturalist's Journal*, **12**: 259-260.

Hamilton PM & Tyler C (2008). Identification of microsatellite loci for parentage analysis in roach *Rutilus rutilus* and 8 other cyprinids by cross-species amplification, and a novel test for detecting hybrids between roach and other cyprinids. *Molecular Ecology Resources*, **8**: 462-465

Hänfling B & Brandl R (1998). Genetic differentiation of the bullhead *Cottus gobio* L. across watersheds in central Europe: evidence for two taxa. *Journal of Fish Biology*, **52**: 706-715.

Hänfling B, Durka W and Brandl R (2004). Impact of habitat fragmentation on genetic population structure of roach, *Rutilus rutilus*, in a riparian ecosystem. *Conservation Genetics*, **5**: 247-257.

Hänfling B & Weetman D (2006). Concordant genetic estimators of migration reveal anthropogenically enhanced source-sink population structure in the river sculpin, *Cottus gobio*. *Genetics*, **173**: 1487-1501.

Hänfling B, Dümpelmann, Bogutskaya NG, Brandl R and Brändle (2009). Shallow phylogeographic structuring of *Vimba vimba* across Europe suggests two distinct refugia during the last glaciation. *Journal of Fish Biology*, **75**: 2269-2286.

Hanski I (1999). *Metapopulation Ecology*. Oxford University Press, Oxford.

Hanksi I & Ovaskainen O (2000). The metapopulation capacity of a fragmented landscape. *Nature*, **404**: 755-758.

Hansson B & Westerberg L (2002). On the correlation between heterozygosity and fitness in natural populations. *Molecular Ecology*, **11**: 2467-2474.

Haplotype Viewer. *http://www.cibiv.at/~greg/haploviewer*.

Hardy OJ. & Vekemans X (2002). SPAGeDi: a versatile computer program to analyze spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**(4): 618-620.

Hardy OJ, Charbonnel N, Fréville H and Heuertz M (2003). Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. *Genetics*, **163**: 1467-1482.

Harland J (2007). Technical Report: The fish bone from Grand Arcade, Cambridge (site code: GAD05/06). *Reports from the Centre for Human Palaeoecology, University of York*, **2008/05**.

Härmä M, Lappalainen A and Urho L (2008). Reproduction areas of roach (*Rutilus rutilus*) in the northern Baltic Sea: potential effects of climate change. *Canadian Journal of Fisheries and Aquatic Sciences*, **65**(12): 2678-2688.

Harpending H (1994). Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology*, **66**: 591-600.

Harris CA, Hamilton PB, Runnalls TJ, Vinciotti V, Henshaw A, Hodgson D, Coe TS, Jobling S, Tyler CR and Sumpter JP (2011). The consequences of feminization in breeding groups of wild fish. *Environmental Health Perspectives*, **119**(3): 306-311.

Harris GS (1978). *Salmon Propagation in England and Wales*. National Water Council, London.

Hasegawa M, Kishino H and Yano T (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, **22**: 160-174.

Hastings A (1993). Complex interactions between dispersal and dynamics – lessons from coupled logistic equations. *Ecology*, **74**: 1362-1372.

Hauser L & Carvalho GR (2008). Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. *Fish and Fisheries*, **9**(4): 333-362.

Hayden B, Coscia I and Mariani S (2011). Cytochrome b variation in bream *Abramis brama*. *Journal of Fish Biology*, **78**: 1579-1587.

Hays JD, Imbrie J and Shackleton NJ (1976). Variation in earth's orbit: pacemaker on the ice ages. *Science*, **194**: 1121-1132.

He S, Mayden RL, Wang, X, Wabg, W, Tang KL, Chen W-J, Chen, Y. (2008). Molecular phylogenetics of the family Cyprinidae (Actinopterygii: Cypriniformes) as evidenced by sequence variation in the first intron of S7 ribosomal protein-coding gene: Further evidence from a nuclear gene of the systematic chaos in the family. *Molecular Phylogenetics and Evolution*, **46**: 818-829.

Hedrick PW (1986). Genetic polymorphism in heterogeneous environments: a decade later. *Annual Review of Ecology and Systematics*, **17**: 535-566.

Hedrick PW (2005a). *Genetics of Populations*, 3rd edition. Jones & Bartlett Publishers, Sudbury.

Hedrick PW (2005b). A standardized genetic differentiation measure. *Evolution*, **59**: 1633-1638.

Hedrick PW, Ginevan ME and Ewing EP (1976). Genetic polymorphism in heterogeneous environments. *Annual Review of Ecology and Systematics*, **7**: 1-32.

Heller R & Siegismund HR (2009). Relationship between three measures of genetic differentiation G_{ST}, D_{EST} and G'_{ST}: how wrong have we been? *Molecular Ecology*, **18**(10): 2080-2083.

Hemming SR (2004). Heinrich events: Massive late Pleistocene detritus layers of the North Atlantic and their global climate imprint. *Reviews of Geophysics*, **42**(1): RG1005.

Hernandez-Martich JD & Smith MH (1997). Downstream gene flow and genetic structure of *Gambusia holbrooki* (eastern mosquitofish) populations. *Heredity*, **79**: 295-301.

Hewlett N, Snow J and Britton JR (2009). The role of management practices in fish kills in recreational lake fisheries in England and Wales. *Fisheries Management and Ecology*, **16**(3): 248-254.

Hickley P (1994). Stocking and introduction of fish. In: *Rehabilitation of Freshwater Fisheries*. Fishing News Books, Blackwell Science Ltd., Oxford.

Hickley P (1995). Recreational Fishing in England and Wales. *Report of the Workshop on Recreational Fishery Management and Planning*, Food and Agriculture Division of the United Nations.

Hickley P & Chare S (2004). Fisheries for non-native species in England and Wales: angling or the environment? *Fisheries Management and Ecology*, **11**: 203-212.

Hill WG (1981). Estimation of effective population size from data on linkage disequilibrium. *Genetical Research*, **38**: 209-216.

Hoarau G, Holla S, Lescasse R, Stam WT and Olsen JL (2002). Heteroplasmy and evidence for recombination in the mitochondrial control region of the flatfish *Platichthys flesus*. *Molecular Biology and Evolution*, **19**: 2261-2264.

Hodder KH & Bullock JM (1997). Translocations of native species in the UK: implications for biodiversity. *Journal of Applied Ecology*, **34**(3): 547-565.

Hoehn M, Sarre SD and K. Henle (2007). The tales of two geckos: does dispersal prevent extinction in recently fragmented populations? *Molecular Ecology*, **16**: 3299–3312

Holm S (1979). A simple sequentially rejective multiple test procedure. *Scandinavian. Journal of. Statistics*, **6**: 65-70.

Holmen J, Vøllestad LA, Jakobsen KS and Primmer CR (2005). Cross-species amplification of zebrafish and stoneroller microsatellite loci in six other cyprinids. *Journal of Fish Biology*, **66**: 851-859.

Holsinger KE & Weir BS (2009). Genetics in geographically structured populations: defining, estimating and interpreting F_{ST}. *Nature Reviews Genetics*, **10**: 639-650.

Hubisz M, Falush D, Stephens M and Pritchard J (2009). Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**: 1322-1332.

Hudson RR, Slatkin M and Maddison WP (1992). Estimation of levels of gene flow from DNA sequence data. *Genetics*, **132**: 583-589.

Huey JA, Baker AM and Hughes JM (2010). High levels of genetic structure in the Australian freshwater fish, *Ambassis macleayi*. *Freshwater Science*, **29**(3): 1148-1160.

Hughes GJ, Fearnhill E, Dunn D, Lycett SJ, Rambaut A and Leigh Brown AJ (2009). Molecular phylodynamics of the heterosexual HIV epidemic in the United Kingdom. *PLoS Pathogens*, **5**(9): e1000590. doi:10.1371/journal.ppat.1000590.

Hughes JM, Ponniah M, Hurwood DA, Chenoweth SF and Arthington A (1999). Strong genetic structuring in a habitat specialist, the oxleyean pygmy perch *Nannoperca oxleyana*. *Heredity*, **83**: 5-14.

Hughes SN & Willis DJ (2000). *The fish community of the River Thames: status, pressures and management*. In: Management and Ecology of River Fisheries (ed – Cowx I G). Fishing News Books, Blackwell Science.

Hutchinson DW & Templeton AR (1999). Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution*, **53**: 1898-1914.

Iervolino F, Kawakami de Resende E, Wagner Silva Hilsdorf A (2010). The lack of genetic differentiation of pacu (*Piaractus mesopotamicus*) populations in the Upper-Paraguay Basin revealed by the mitochondrial DNA D-loop region: implications for fishery management. *Fisheries Research*, **101**: 27-31.

Ihssen PE, Booke HE, Casselman JM, McGlade JM, Payne NR and Utter FM (1981). Stock identification: materials and methods. *Canadian Journal of Fisheries and Aquatic Sciences*, **38**: 1838-1855.

Imsiridou A, Karakoussis Y and Triantafyllidis C (1997). Genetic polymorphism and differentiation among chub *Leuciscus cephalus* (Pisces, Cyprinidae) populations of Greece. *Biochemical Systematics and Ecology,* **25**: 537-546.

Imsiridou A, Apostolidis A, Durand J, Briolay J, Bouvet Y and Triantafyllidis C (1998). Genetic differentiation and phylogenetic relationships among Greek chub L. *Leuciscus cephalus* (Pisces, Cyprinidae) populations as revealed by RFLP analysis of amplified mitochondrial DNA segments. *Biochemical Systematics and Ecology*, **26**: 415-429.

Ingman M & Gyllensten U (2007). Rate variation between mitochondrial domains and adaptive evolution in humans. *Human Molecular Genetics*, **16**: 2281-2287.

Ingvarsson PK, Olsson K and Ericson K (1997). Extinction-recolonisation dynamics in the mycophagous beetle *Phalacrus substriatus*. *Evolution*, **51**: 187-195.

Jafri SIH (1987). Incidental breeding of roach (*Rutilus rutilus*) in an Aquarium. *The Progressive Fish Culturalist*, **49**: 237-238.

Jafri SH & Ensor DM (1976). Occurrence of an intersex condition in the roach *Rutilus rutilus* (L). *Journal of Fish Biology*, **14**(6): 547-549.

Jean C-T, Hui C-F, Lee S-C and Chen C-T (1995). Variation in mitochondrial DNA and phylogenetic relationships of fishes of the sudfamily Sparinae (Perciformes: Sparidae) in the coastal waters of Taiwan. *Zoological Studies*, **34**: 270-280.

Jeffreys AJ, Wilson V and Thein SL (1985). Hypervariable 'minisatellite' regions in human DNA. *Nature*, **314**: 67-73.

Jeffreys H (1967). *Theory of probability*. Clarendon Press, Claredon.

Jenkins DG, Brescacin CR, Duxbury CV, Elliott JA, Evans JA, Grablow KR, Hillegass M, Lyon BN, Metzger GA, Olandese ML, Pepe D, Silvers GA, Suresch HN, Thompson TN, Trexler CM, Williams GE, Williams NC and Williams SE (2007). Does size matter for dispersal distance? *Global Ecology and Biogeography*, **16**: 415-425.

Jensen JL Bohonak AJ and Kelley ST (2005). Isolation by distances, web service. *BMC Genetics*, **6**: 13.

Jepsen N & Berg S (2002). The use of winter refuges by roach tagged with miniature radio transmitters. *Hydrobiologia*, **483**: 167-173.

Jobling S, Coey S, Whtitmore JG, Kime DEvan Look KJW, McAllister BG, Beresford N, henshaw AC, Bright G, Tyler CR and Sumpter JP (2002). Wild intersex roach (*Rutilus rutilus*) have reduced fertility. *Biological Reproduction*, **67**: 515-524.

Johannesson K & André C (2005). Life on the margin: genetic isolation and diversity loss in a peripheral marine system, the Baltic Sea. *Molecular Ecology*, **15**: 2013-2029.

Johnson AC, Acreman MC, Dunbar MJ, Feist SW, Giacomello AM, Gozlan RE, Hinsley SA, Ibbotson AT, Jarvie HP, Jones JI, Longshaw M, Maberly SC, Marsh TJ, Neal C, Newman JR, Nunn MA, Pickup RW, Reynard NS, Sullivan CA, Sumpter JP and Williams RJ (2009). The British river of the future: How climate change and human activity might affect two contrasting river ecosystems in England. *Science of the Total Environment*, **407**: 4787-4798.

Jones O & Wang J (2009). COLONY: a program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources*, **10**: 551-555.

Jorde P E & Ryman N (1995). Temporal allele frequency change and estimation of effective size in populations with overlapping generation. *Genetics*, **139**: 1077-1090.

Jost L (2008). G_{ST} and its relatives do not measure differentiation. *Molecular Ecology*, **17**: 4015-4026.

Jost L (2009). D vs. G_{ST}: Response to Ryman and Leimer. *Molecular Ecology*, **18**(10): 2088-2089.

Jukes TH & Cantor CR (1969). Evolution of protein molecules. In: *Mammalian Protein Metabolism* (ed – Munro HM). Academic, New York.

Karlsen HE, Piddlington RW, Enger PS and Sand O (2004). Infrasound initiates directional fast-start escape responses in juvenile roach (*Rutilus rutilus). Journal of Experimental Biology*, **207**: 4185-4193.

Kawamura K, Ueda T and Arai R (2001). Genetic introgression by the rose bitterling, *Rhodeus ocellatus ocellatus*, into the Japanese rose bitterling, *Rhodeus ocellatus kurumeus* (Teleostei: Cyprinidae). *Zoological Science*, **18**: 1027-1039.

Kelly C & Rice J (1996). Modelling nucleotide evolution: a heterogeneous rate analysis. *Mathematical Biosciences*, **133**(1): 85-109.

Kennedy M and Fitzmaurice P (1973). Occurrence of cyprinid hybrids in Ireland. *Irish Naturalist's Journal*, **17**: 349-351.

Ketmaier V, Bianco PG and Durand J-D (2008). Molecular systematics, phylogeny and biogeography of roaches (Rutilus, Teleostei, Cyprinidae). *Molecular Phylogenetics and Evolution*, **49**(1): 362-367.

Kimura M (1968). Evolutionary rate at the molecular level. *Nature*, **217**: 624-626.

Kimura M (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**: 111-120.

Kimura M (1983). *The Neutral Theory of Molecular Evolution*. Cambridge University Press, UK.

Kimura M & Crow J (1964). The number of alleles that can be maintained in a finite population. *Genetics*, **49**: 725-738.

Kimura M & Weiss GH (1964). The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics*, **49**: 561-576.

Kimura M & Ohta T (1973). Mutation and evolution at the molecular level. *Genetics*, **73**: 19-35.

Kimura M & Ohta T (1978). Stepwise mutation model and distribution of allele frequencies in a finite population. *Proceedings of the National Academy of Sciences USA*, **75**: 2868-2872.

King JL & Jukes TH (1969) Non-Darwinian evolution. *Science*, **164**: 788-798.

Kingman JFC (1982a). On the genealogy of large populations. *Journal of Applied Probability*, **19A**: 27-43.

Kingman JFC (1982b). The coalescent. *Stochastic Process Applications*, **13**: 235-248.

Knaepkens G, Baekelandt K and Eens M (2006). Fish pass effectiveness for bullhead (*Cottus gobio*), perch (*Perca fluviatilis*) and roach (*Rutilus rutilus*) in a regulated lowland river. *Ecology of Freshwater Fish*, **15**(1): 20-29.

Knapen D, Knaepkens G, Bervoets L, Taylor MI, Eens M and Verheyen E (2003). Conservation units based on mitochondrial and nuclear DNA variation among European bullhead (*Cottus gobio*) populations: implications for conservation. *Biological Conservation*, **115**: 403-410.

Kohlmann K, Kersten P and Flajšhans (2007). Comparison of microsatellite variability in wild and cultured tench (*Tinca tinca*). *Aquaculture*, **272** (Supplement 1): S147-S151.

Kokko H & López-Sepulchre A (2006). From individual dispersal to species range: perspectives for a changing world. *Science*, **313**: 789-791.

Kontula T & Väinölä R (2001). Postglacial colonization of Northern Europe by distinct phylogeographic lineages of the bullhead, *Cottus gobio*. *Molecular Ecology*, **10**(8): 1983-2002.

Kortet R, Vainikka A, Rantala MJ, Jokinen I and Taskinen J (2003). Sexual ornamentation, androgens and papillomatosis in male roach (*Rutilus rutilus*), *Evolutionary Ecology Research*, **5**: 411-419.

Kortet R, Taskinen J, Vainikka A and Ylönen H (2004). Breeding tubercles, papillomatosis and dominance behaviour of male roach (*Rutilus rutilus*) during the spawning period. *Ethology*, **110**(8): 591-601.

Koskinen MT, Ranta E, Piironen J, Veselov A, Titov S, Haugen TO, Nilsson J, Carlstein M and Primmer CR (2000). Genetic lineages and postglacial colonization of grayling (*Thymallus thymallus*, Salmonidae) in Europe, as revealed by mitochondrial DNA analyses. *Molecular Ecology*, **9**(10): 1609-1624.

Koskinen MT, Nilsson J, Veselov AJ, Potutkin AG, Ranta E and Primmer CR (2002). Microsatellite data resolve phylogeographic patterns in European grayling, *Thymallus thymallus*, Salmonidae. *Heredity*, **88**: 391 -401.

Koskinen MT, Hirvonen H, Landry P-A and Primmer CR (2004). The benefits of increasing the number of microsatellites utilised in genetic population studies: an empirical perspective. *Hereditas*, **141**: 61-67.

Kottelat M & Freyhof J (2007). *Handbook of European Freshwater Fishes*. Kottelat, Cornol, Switzerland and Freyhof, Berlin, Germany.

Krieger J & Fuerst PA (2004). Evidence of multiple alleles of the nuclear 18S ribosomal RNA gene in sturgeon (Family: Acipenseridae). *Journal of Applied Ichthyology*, **18**: 290-297.

Kruse CG, Hubert WA and Rahel FJ (1997). Geomorphic influences on the distribution of Yellowstone cutthroat trout in Absaroka Mountains, Wyoming. *Transactions of the American Fisheries Society*, **126**: 418-427.

L'Abée-Lund JH & Vøllestad LA (1985). Homing precision of roach *Rutilus rutilus* in Lake Årungen, Norway. *Environmental Biology of Fishes*, **13**(3): 235-239.

Lajbner Z, Linhart O and Kotlík P (2011). Human-aided dispersal has altered the phylogeography of the tench (*Tinca tinca*). *Evolutionary Applications* (doi: 10.1111/j.1752-4571.2010.00174.x).

Lande R (1988). Genetics and demography in biological conservation. *Science*, **241**: 1455–1460.

Lande R (1994). Risk of population extinction from fixation of new deleterious mutations. *Evolution*, **48**: 1460-1469.

Lande R & Barrowclough GF (1987). Effective population size, genetic variation, and their use in population management. In*: Viable Populations for Conservation* (ed – Soulé M). Cambridge University Press. Cambridge, England.

Lanters RLP (1993). De bekkenvistrap Belfeld: Monitoring van de visoptrek en hydraulische waarnemingen in 1993. Rapport 93. 023 RIVO-DLO, Ijmuisden.

Lanters RLP (1995). Vismigratie door de bekkenvistrappen lith en Belfeld de

Larinier M (2001). Environmental issues, dams and fish migration. In: *Dams, Fish and Fisheries: opportunities, challenges and conflict resolution*. FAO, Fisheries Technical Paper, United Nations.

Larmuseau MHD, Freyhof J, Volkaert FAM and Van Houdt JKJ (2009). Matrilineal phylogeography and demographical patterns of *Rutilus rutilus*: implications for taxonomy and conservation. *Journal of Fish Biology*, **75**: 332-353.

Laroche J, Durand JD, Bouvet Y, Guinand B and Brohon B (1999). Genetic structure and differentiation among populations of two cyprinids, *Leuciscus cephalus* and *Rutilus rutilus*, in a large European River. *Canadian Journal of Fisheries and Aquatic Science*, **56**: 1659-1667.

Le Rouzic & Carlborg O (2008). Evolutionary potential of hidden genetic variation. *Trends in Ecology and Evolution*, **23**(1): 33-37.

Lee W, Conroy J, Howell WH and Kocher TD (1995). Structure and evolution of teleost mitochondrial control regions. *Journal of Molecular Evolution*, **41**: 54–66.

Levene H (1949). On a matching problem arising in genetics. *Annals of Mathematical Statistics*, **20**: 91-94.

Levene H (1953). Genetic equilibrium when more than one ecological niche is available. *American Naturalist*, **87**: 331-333.

Levene H (1960). Robust tests for equality of variances. In: *Contributions to Probability and Statistics* (ed – Olkin I). Stanford University Press, USA.

Levins R (1969). Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the Entomological Society of America*, **15**: 237-240.

Levins R (1970). Extinction. In: *Some Mathematical Problems in Biology* (ed – Gerstenhaber M). American Mathematical Society, Providence, USA.

Lewontin RC & Kojima K (1960). The evolutionary dynamics of complex polymorphisms. *Evolution*, **14**: 450-472.

Lewontin RC & Felsenstein J (1965). The robustness of homogeneity tests in 2 x N tables. *Biometrics*, **21**: 19-33.

Lewontin RC & Krakauer J (1973). Distribution of gene frequency as a test of the theory of selective neutrality of polymorphisms. *Genetics*, **75**: 175-195.

Li W-H (1977). Distribution of two randomly chosen cistrons in a finite population. *Genetics*, **85**: 331-337.

Librado P & Rozas J (2009). DNAsp v.5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**: 1451-1452.

Liu G, Zhou J and Zhou D (2012). Mitochondrial DNA reveals low population differentiation in elongate loach, *Leptobotia elongata* (Bleeker): implications for conservation. *Environmental Biology of Fishes*, **93**(3): 393-402.

Liu H (2002). Mitochondrial DNA sequence variations and systematics of the genus *Distoechodon* (Teleostei: Cyprinidae). *Journal of Applied Ichthyology*, **18**: 181-184.

Lopes-Cunha M, Aboim MA, Mesquita N, Alves JM, Doadrio I and Coelho MM (2012). Population genetic structure in the Iberian cyprinid fish, *Iberochondrostoma lemmingii* (Steindachner, 1866): disentangling species fragmentation and colonization processes. *Biological Journal of the Linnaean Society*, **105**(3): 559-572.

Luca C, Manea A, Luca A, Dinischiotu A and Costache M (2008). Molecular phylogeny of the Romanian cyprinids from the Danube River. *Romanian Biotechnological Letters*, **13**(5): 3970-3975.

Lucas MC & Frear PA (1997). Effects of a flow-gauging weir on the migratory behavior of adult barbell, a riverine cyprinid. *Journal of Fish Biology*, **50**(2): 382-396.

Lucas MC and Baras E (2001). *Migration of Freshwater Fishes*. Blackwell Science Ltd., Oxford.

Ludwig A, May B, Debus L and Jenneckens I (2000). Heteroplasmy in the mtDNA control region of sturgeon (*Acipenser, Huso and Scaphirynchus*). *Genetics*, **156**: 1933-1947.

Luikart G, Allendorf FW, Cornuet JM and Sherwin WB (1998). Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity*, **89**: 238-247.

Luikart G & Cornuet JM (1998). Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conservation Biology*, **12**(1): 228-237.

Luikart GR, Cornuet J-M and Allendorf FW (1999). Temporal changes in allele frequencies provide estimates of population bottleneck size. *Conservation Biology*, **13**(3): 523-530.

Lynch M, Conery JS and Burger R (1995). Mutation accumulation and the extinction of small populations. *American Naturalist*, **146**: 489-518.

Lynch M & Conery JS (2003). The origins of genomic complexity. *Science*, **302**: 1401-1404.

Lyons J, Hickley P and Gledhill S (2002). An evaluation of recreational fishing in England and Wales. In: *Evaluating Recreational Fishing* (eds – Pitcher TJ & Hollingworth CE). Blackwell Science Ltd., Oxford.

MacArthur RH & Wilson EO (1967). *The Theory of Island Biogeography*. Princeton University Press, USA.

Mace AE (1964). *Sample-Size Determination*. Reinhold, New York, USA.

Maes J, Breine J, Stevens, M and Ollevier (2005). New perspectives for fish in the Scheldt estuary. In: *Proceedings 'Dunes and Estuaries 2005' – International Conference in Nature Restoration* (eds – Herrier J-L, Mees J, Salman A, Seys H, Van Nieuwenhuyse H, and Dobbelaere L). VLZ Special Publication 16.

Magoulas A & Zouros E (1993). Restriction-site heteroplasmy in anchovy (*Engraulis encrasicolus*) indicates incidental biparental inheritance of mitochondrial DNA. *Molecular Biology and Evolution*, **10**: 319-325.

Maitland PS (1979). The status and conservation of rare freshwater fishes in the British Isles. *Proceedings of the Freshwater Fisheries Conference*, **12**: 265-278.

Maitland PS (1986). The potential impact of fish culture on wild stocks of Atlantic salmon in Scotland. In: *The Status of the Atlantic Salmon in Scotland* (eds – Jenkins D & Shearer WM). ITE Symposium No. 15, HMSO, London.

Maitland PS (1987). Fish introductions and translocations – their impact in the British Isles. *Institute of Terrestrial Ecology Symposium*, **19**: 57-65

Maitland PS (1995). The conservation of freshwater fish: past and present experience. *Biological Conservation*, **72**: 259-270.

Maitland PS and Linsell K (2006). *Philip's Guide to Freshwater Fish of Britain and Europe*. Octopus Publishing Group Ltd., London.

Malécot G (1948). *Les mathématiques de l'héréditié*. Masson et Cie, Paris.

Malhotra M & Thorpe RS (1994). Parallels between island lizards suggests selection on mitochondrial DNA and morphology. *Proceedings of the Royal* Society *B*, **257**: 37-42.

Mamirova L, Popadin and Gelfand MS (2007). Purifying selection in mitochondria, free-living and obligate intracellular protobacteria. *BMC Evolutionary Biology*, **7**: 17.

Mann RHK (1996). Environmental requirements of European non-salmonid fish in rivers. *Hydriobiologia*, **323**: 223-235.

Manni F, Guérard and Heyer E (2004). Geographic patterns of (genetic, morphologic, linguistic) varation: how barriers can be detected by using Monmonier's algorithm. *Human Biology*, **76**(2): 173-190.

Mantel N (1967). The detection of disease clustering and a generalized regression approach". *Cancer Research*, **27**(2): 209–220.

Marchetto F, Zaccara, Muenzel FM and Salzburger W (2010). Phylogeography of the Italian vairone (*Telestes muticellus*, Bonaparte 1837) inferred by microsatellite markers: evolutionary history of a freshwater fish species with a restricted and fragmented distribution. *BMC Evolutionary Biology*, **10**: 111.

Marsh TJ and Lees ML (2003). *Hydrometric register and statistics 1996-2000, hydrological data UK series*. CEH Wallingford/British Geological Survey, Wallingford, UK.

Martin AP, Naylor G and Palumbi S (1992). Rates of mitochondrial DNA evolution in sharks are slow compared with mammals. *Nature*, **357**: 153-155.

Maruyama T & Kimura M (1980). Genetic variability and effective population size when local extinction and recolonization of subpopulations are frequent. *Proceedings of the National Academy of Sciences USA*, **77**: 6710-6714.

Matthews CP (1971). Contribution of young fish to total production of fish in the River Thames near Reading. *Journal of Fish Biology*, **3**: 157-180.

Matthews WJ (1998). *Patterns in Freshwater Fish Ecology*. Chapman and Hall, New York.

Mawle GW & Milner NJ (2003). The return of salmon to cleaner rivers – England and Wales. In: *Salmon at the Edge* (ed – Mills D). Blackwell Science Ltd.

Maxim A and Gilbert W (1977). A new method for sequencing DNA. *Proceedings of the. National. Academy of. Sciences USA*, **74**(2): 560-564.

Mayr E (1940). Speciation phenomena in birds. *American Naturalist*, **74**: 249-278.

Mayr E (1954). Change in the genetic environment and evolution. In: *Evolution as a Process* (ed – Huxley J). Allen & Unwin, London.

McCusker MR & Bentzen P (2010). Positive relationship between genetic diversity and abundance in fishes. *Molecular Ecology*, **19**(22): 4852-4862.

McGlashan DJ & Hughes JM (2001). Low levels of genetic differentiation among populations of the freshwater fish *Hypseleotris compressa* (Gobiidae: Eleotinidae): implications for its biology, population connectivity and history. *Heredity*, **86**: 222-233.

Meirmans PG (2006). Using the AMOVA framework to estimate a standardized genetic differentiation measure. *Evolution*, **60**(11): 2399-2402.

Meirmans PG & van Tienderen PH (2004). GENETYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Resources*, **4**(4): 792-794.

Meirmans PG & Hedrick PW (2011). Assessing population structure: F_{ST} and related measures. *Molecular Ecology Notes*, **11**(1): 5-18.

Mercer JP (1968). Roach, *Rutilus rutilus* L. – a new record from Co. Fermanagh. *Irish Naturalist's Journal*, **16**: 52.

Meselson M and Yuan R (1968). DNA restriction enzyme from *E. coli*. *Nature*, **217**: 1110-1114.

Mesquita N, Carvalho G, Shaw P, Crespo E and Coelho MM (2001). River basin-related genetic structuring in an endangered fish species, *Chondrostoma lusitanicum*, based on mtDNA sequencing and RFLP analysis. *Heredity*, **86**: 253-264.

Mesquita N, Hänfling B, Carvalho GR and Coehlo MM (2005). Phylogeography of the cyprinid *Squalius aradensis* and implications for conservation of the endemic freshwater fauna of southern Portugal. *Molecular Ecology*, **14**: 1939-1954.

Meyer A (1993). Evolution of mitochondrial DNA in fishes. *Biochemistry and Molecular Biology of Fishes*, Vol 2. Elsevier Press, Netherlands.

Michalakis Y & Excoffier L (1996). A generic estimation of population subdivisions using distances between alleles with special reference to microsatellite loci. *Genetics*, **142**: 1061-1064.

Mighell AJ, Smith NR, Robinson PA and Markham AF (2000). Vertebrate pseudogenes. *FEBS Letters*, **468**: 109-114.

Millar CI & Libby WJ (1994). Disneyland or native ecosystem: genetics and the restorationist. *Restoration & Management Notes*, **7**: 18-24.

Mills CA (1981). The spawning of Roach *Rutilus rutilus* (L.) in a chalk stream. *Aquaculture Research*, **12**(2): 49.54.

Mitton JB & Lewis WM (1989). Relationships between genetic variability and life-history features of bony fishes. *Evolution*, **43**: 1712-1723.

Miyazono F, Schneider PM, Metzger R, Warnecke-Eberz U, Baldus SE, Dienes HP, Aikou T and Hoelscher AH (2002). Mutations in the mitochondrial DNA D-Loop occur frequently in adenocarcinoma in Barrett's oesophagous. *Oncogene*, **21**: 3780-3783.

Mjelle KA, Karlsen BO, Jorgensen TE, Moum T and Johansen SD (2008). Halibut mitochondrial genomes contain extensive heteroplasmic tandem repeat arrays involved in DNA recombination. *BMC Genomics*, **9**: 10.

Moon N & Souter G (1994) Soci-economic review of angling 1994. *Fisheries Technical Report* No. 5. Bristol: National Rivers Authority.

Moore RV, Morris DG and Flavin RW (1994). Sub-set of UK digital 1:50000 scale river centre-line network. NERC Institute of Hydrology, Wallingford.

Moran MD (2003). Arguments for rejecting the sequential Bonferroni in ecological studies. *Oikos*, 100: 403-405.

Morin PA, Archer FI, Foote AD, Vilstrup J, Allen EE, Wade P, Durban J, Parsons K, Pitman R, Li L, Bouffard P, Abel Nielsen SC, Rasmussen M, Willerslev E, Gilbert MT and Harkins T (2010). Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. *Genome Research*, **20**(7): 908-916.

Moritz C (1994). Defining 'Evolutionary Significant Units' for conservation. *Trends in Ecology and Evolution*, **9**: 373-375.

Moritz C (2002). Strategies to protect biological diversity and the processes that sustain it. *Systematic Biology*, 51: 238-254.

Moritz C, Dowling TE and Brown WM (1987). Evolution of animal mitochondrial DNA: relerence for population biology and systematic. *Annual Review of Ecology and Systematics*, **18**: 269-292.

Moyer GR, Winemiller KO, McPhee MV and Turner TF (2005). Historical demography, selection, and coalescence of mitochondrial and nuclear genes in *Prochilodus* species of northern South America. *Evolution*, **59**: 599-610.

Müller K (1954). Investigations on the organic drift in North Swedish streams. *Institute of Freshwater Research Reports*, **34**: 133-148.

Nagylaki T (1978). Clines with asymmetric migration. *Genetics*, **88**: 813-827.

Nagylaki T (1998). Fixation indices in subdivided populations. *Genetics*, **148**: 1325-1332.

Nakagawa S (2004). A farewell to Bonferroni: the problems of low statistical power and publication bias. *Behavioural Ecology*, **15**(6): 1044-1045.

Narum SR (2006). Beyond Bonferroni: less conservative analyses for conservation genetics. *Conservation Genetics*, **7**(5): 783-787.

National Rivers Authority (1995) National Angling Survey 1994. NRA *Fisheries Technical Report* No.5.

Nee S, Holmes EC, Rambaut A and Harvey PH (1996). Inferring population history from molecular phylogenies. In: *New Uses for New Phylogenies* (eds – Harvey PH, Leigh Brown AJ, Maynard Smith J and Nee S). Oxford University Press, UK.

Nei M (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences USA*, **70**: 3321-3323.

Nei M (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**: 583-590.

Nei M (1987). *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nei M & Li W-H (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences USA*, **76**: 5269-5273.

Nei M & Tajima F (1984). Note on genetic drift and estimation of effective population size. *Genetics*, **106**: 569-574.

Nei M & Kumar S (2000). *Molecular Evolution and Phylogenetics*. Oxford University Press, Oxford.

Neigel JE (1997). A comparison of alternative strategies for estimating gene flow from genetic markers. *Annual Review of Ecology, Evolution and Systematics*, **28**: 105-128.

Neigel JE (2002). Is F_{ST} obsolete? *Conservation Genetics*, **3**: 167-173.

Nelson JS (2006). *Fishes of the World*. Wiley & Son, Inc., New Jersey.

Nesbø CL, Fossheim T, Vøllestad LA and Jakobsen KS (1999). Genetic divergence and phylogeographic relationships among European perch (*Perca fluviatilis*) populations reflect glacial refugia and postglacial colonization. *Molecular Ecology*, **8**: 1387-1404.

Network version 4.6. *www.flexus-engineering.com*

Newton MA & Rafferty AE (1994). Approximate Bayesian inference by the weighted likelihood bootstrap. *Journal of the Royal Statistical Society Series B*, **56**: 3-48.

Nielsen EE, Hansen MM and Loeshke V (1996). Genetic structure of European populations of *Salmo salar* L. (Atlantic salmon) inferred from mitochondrial DNA. *Heredity*, **77**: 351-358.

Nielsen R & Wakeley J (2001). Distinguishing migration from isolation: a Markov Chain Monte Carlo approach. *Genetics*, **158**(2): 885-896.

Nilsson J, Gross R, Asplund T, Dove O, Jansson H, Kelloniemi J, Kohlmann K, Löytynoja, Nielsen EE, Paaver T, Primmer CR, Titov S, Vasemägi A, Veselov A, Östs T and Lumme J (2001). Matrilinear phylogeography of Atlantic salmon (*Salmon salar* L.) in Europe and postglacial colonization of the Baltic Sea area. *Molecular Ecology*, **10**: 89-102.

Ning T, Xiao H, Li J, Hua S and Zhang YP (2010). Adaptive evolution of the mitochondrial ND6 gene in the domestic horse. *Genetics and Molecular Research*, **9**(1): 144-150.

Nixon KC & Wheeler QD (1990). An amplification of the phylogenetic species concept. *Cladistics*, **6**: 211-223.

Nõges P & Järvet A (2005). Climate driven changes in the spawning of roach (*Rutilus rutilus* (L.)) and bream (*Abramis brama* (L.)) in the Estonian part of then Narva River basin. *Boreal Environment Research*, **10**: 45-55.

Nolan M, Jobling S, Brighty G, Sumpter JP and Tyler CR (2001). A histological description of intersexuality in the roach. *Journal of Fish Biology*, **58**: 160-176.

Nordborg M & Krone SM (2002). Separation of time scales and convergence to the coalescent in structured populations. In: *Modern Developments in Theoretical Population Genetics: The Legacy of Gustave Malécot* (eds – Slatkin M & Veuille M). Oxford University Press, Oxford.

Nunney L (1995). Measuring the ratio of effective population size to adult numbers using genetic and ecological data. *Evolution*, **49**(2): 389-392.

O'Grady JJ, Brook BW, Reed DH, Ballou JD, Tonkyn DW and Frankham R (2006). Realistic levels of inbreeding depression strongly affect extinction risk in wild populations. *Biological Conservation*, **133**: 42-51.

Ohta T (1973). Slightly deleterious mutant substitution in evolution. *Nature*, **246**: 96-98.

Ohta T (1992). The nearly neutral theory of molecular evolution. *Annual Review of Ecological Systematics*, **23**: 263-286.

Osinov AG & Bernatchez L (1996). "Atlantic" and "Danubian" phylogenetic groupings of brown trout *Salmo trutta* complex: genetic divergence, evolution and conservation. *Journal of Ichthyology*, **36**: 723-746.

Paetkau D, Calvert W, Stirling I and Strobeck C (1995). Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology*, **4**:347–354.

Paetkau D & Strobeck C (1995). The molecular basis and evolutionary history of a microsatellite null allele in bears. *Molecular Ecology*, **4**: 519-520

Palsbøl PJ, Bérubé M and Allendorf W (2006). Identification of management units using population genetic data. *Trends in Ecology and Evolution*, **22**(1): 11-16.

Palsbøl PJ, Zachariah Peery M and Bérubé (2010). Detecting populations in the 'ambiguous' zone: kinship-based estimation of population structure at low genetic divergence. *Molecular Ecology Resources*, **10**(5): 797-805.

Palstra FP & Ruzzante DE (2008). Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild population persistence. *Molecular Ecology*, **17**: 3428-3447.

Panchal M (2007). The automation of Nested Clade Phylogeographic Analysis. *Bioinformatics Applications*, **23**(4): 509-510.

Pannell JR & Charlesworth B (1999). Neutral genetic diversity in a metapopulation with recurrent local extinction and recolonization. *Evolution*, **53**: 664-676.

Pattee E (1988). Fish and their environment in large European river systems. The Rhone. *Science L'Eau*, **7**: 35-74.

Peakall R & Smouse PE (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**: 288-295.

Peel D, Ovenden JR & Peel SL (2004). NeEstimator: software for estimating effective population size. Version 1.3. Queensland Government, Department of Primary Industries and Fisheries.

Peichel CL, Nereng KS, Ohgi KA, Cole BL, Colosimo PF, Buerkle GA, Schlutter D and Kingsley DM (2001). The master sex-determination locus in three-spine sticklebacks is on a nascent Y chromosome. *Current Biology*, **14**: 1416-1424.

Peirson G, Tingley D, Spurgeon J and Radford A (2001) Economic evaluation of inland fisheries in England and Wales. *Fisheries Management and Ecology,* **8**: 415-424.

Pelz GR (1985). Fischbewegungen über versciedenartige Fischpasse am Beispiel der Mosel. *Courier Forschung Institut Senckenberg* **76**: 1-190.

Pemberton JM, Slate J, Bancroft DR and Barrett JA (1995). Nonamplifying alleles at microsatellite loci: a caution for parentage and population studies. *Molecular Ecology*, **4**: 249-252.

Pena SDJ & Chakraborty R (1994). Paternity testing in the DNA era. *Trends in Genetics*, **10**: 204-209.

Peterka J, Mojmir V, Kubečka J, Hladik M and Hohausová (2004). Drift of juveniles after riverine spawning of fishes from the Římov reservoir, Czech Republic. *Ecohydrology and Hydrobiology*, **4**(4): 459-468.

Pethon P (1980). Variation in the fish community of the 0ra Estuary, SE Norway, with emphasis on the fishes. *Fauna Norv. Ser. A*. **1**: 5-14.

Pfenninger M, Véla E, Jesse R, Elejalde MA, Liberto F, Magnin F and Martínez-Ortí A (2010). Temporal speciation pattern in the western Mediterranean genus *Tudorella* (Fischer 1885) (Gastropoda: Pomatiidae) supports the Tyrrhenian vicariance hypothesis. *Molecular Phylogenetics and Evolution*, **54**(2): 427-436.

Piry S, Luikart G and Cornuet JM (1999). BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity*, **90**: 502-503.

Piry S, Alapetite A, Cornuet JM, Paetkau D, Baudouin L and Estoup A (2004). GENECLASS 2. A software for genetic assignment and first-generation migrant detection. *Journal of Heredity*, **95**: 536-539.

Pitcher TJ & Hollingworth CE (2002). Fishing for fun: where's the catch? In *Recreational Fisheries: Ecological, Economic and Social Evaluation* (eds Pitcher TJ & Hollingworth CE). Blackwell Science Ltd., Oxford.

Pitts CS, Jordan DR, Cowx IG and Jones NV (1997). Controlled breeding studies to verify the identity of roach and common bream hybrids from a natural population. *Journal of Fish Biology*, **51**: 686-696.

Pollak E (1983). A new method for estimating the effective population size from allele frequency changes. *Genetics*, **104**: 531-548.

Posada D (2008). jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, **25**: 1253-1256.

Posada D & Buckley TR (2004). Model selection and model averaging in phylogenetics: advantages of the AIC and Bayesian approaches over likelihood ratio tests. *Systematic Biology*, **53**: 793-808.

Poulet N (2007). Impact of weirs on fish communities in a piedmont stream. *River Research and Applications*, **23**(9): 1038-1047.

Powers DA, Lauerman T, Crawford D, Smith M, Gonzalez-Villasenor I and DiMichele I (1991). The evolutionary significance of genetic variation at enzyme synthesizing loci in the teleost *Fundulus heteroclitus*. *Journal of Fish Biology*, **39** (Supplement A): 169-184.

Pradhi A (2011). Climate-driven population dynamics of a North American freshwater fish, *Ictalurus furcatus*: inference from the mitochondrial control region sequence data. *Ecology of Freshwater Fish*, **20**: 263-269.

Prignon C, Micha JC & Gillet A (1998). Biological and environmental characteristics of fish passage at the Talifer dam on the Meuse River, Belgium. In: *Fish Migration and Fish Bypasses* (eds - Jungworth M, Schmutz S & Weiss S). Fishing News Books, Blackwell Science Ltd., Oxford.

Primmer CR, Rausepp T, Chowdhary BP and Ellegren H (1997). Low frequency of microsatellites in the avian genome. *Genome Research*, **7**(5): 471-482.

Primmer CR, Koskinen MT and Piironen J (2000). The one that did not get away: individual assignment using microsatellite data detects a case of fishing competition fraud. *Molecular Ecology*, **10**: 2375-2388.

Primmer CR, Veselov AJ, Zubichenko APoututkin A, Bakhmeti I and Koskinen MT (2006). population structuring of Atlantic salmon, *Salmo salar*, in tributaries of the Varguza River in northwest Russia. *Molecular Ecology*, **15**: 653-666.

Pringle JM, Blakesbee AMH, Byers JE and Roman J (2011). Asymmetric dispersal allows an upstream region to control population structure throughout a species' range. *Proceedings of the National Academy of Sciences USA*, **108**(37): 15288-15293.

Pritchard JK, Stephens M and Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics*, **155**(2): 945-959.

Provine WB (1987). *The Origin of Theoretical Population Genetics*. University of Chicago Press, Chicago.

Provine WB (1989). *Sewall Wright and Evolutionary Biology*. University of Chicago Press, Chicago.

Pulliam HR (1988). Sources, sinks, and population regulation. *American Naturalist*, **132**: 652-661.

Purdom CE (1993). *Genetics and Fish Breeding*. Chapman and Hall, London.

Radchenko OA (2004). Introgressive hybridization of chars of the genus *Salvelinus* as inferred from mitochondrial DNA variation. *Russian Journal of Genetics*, **40**(12): 1392-1398.

Rambaut A & Drummond AJ (2004). Tracer v1.5.http://evolve.zoo.ox.ac.uk/software.html

Ramos-Onsins SE & Rozas J (2002). Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, **19**: 2092-2100.

Rand DM (1994). Thermal habit, metabolic rate and the evolution of mitochondrial DNA. *Trends in Ecology and Evolution*, **9**: 125-131.

Rannala B & Mountain JL (1997). Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences USA*, **94**: 9197-9201.

Raymond M & Rousset F (1995a). An exact test for population differentiation. *Evolution*, **49**: 1280-1283.

Raymond M & Rousset F (1995b). GENEPOP (Version 1.2) – Population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**: 248-249.

Reed DH (2004). Extinction risk in fragmented habitats. *Animal Conservation*, **7**: 185-193.

Refseth UH, Nesbø CL, Vøllestad LA, Fjeld E, Stacy JE and Jakobsen KS (1998). Genetic evidence for different migration routes of freshwater fish into Norway revealed by analysis of current perch (*Perca fluviatilis*) populations in Scandinavia. *Molecular Ecology*, **7**: 1015-1027.

Rhymer JM & Simberloff D (1996). Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics*, **27**: 83-109.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, **43**: 223-225.

Riggs L (1990). Principles for genetic conservation and production quality. *Report for the Northwest Power Planning Council, Portland, OR, USA* (contract no. C90-005).

Robertson A (1964). The effect of non-random mating within inbred lines on the rate of inbreeding. *Genetics Research*, **5**: 164-167.

Robinson CA, Hickley P and Axford S (2003). The value and performance of large river recreational fisheries in England. *Ecohydrology and Hydrobiology*, **3**(1): 51-60.

Robinson C & Whitton W (2004). Fisheries Action Plans – a new approach to public consultation and the impetus for habitat enhancement works n the upper Thames catchment (England). *3rd European Conference on River Restoration*.

Roe BA, Ma D-P, Wilson RK and Wong F-H (1985). The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome. *Journal of Biological Chemistry*, **260**: 9759-9774.

Rogers AR & Harpending H (1992). Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, **9**: 552-569.

Rogozin IB & Pavlov YI (2003). Theoretical analysis of mutation hotspots and their DNA sequence context specificity. *Mutation Research*, **544**(1): 65-85.

Ronfort J, Jenczcewski E, Bataillon T and Rousset F (1998). Analysis of population structure in autotetraploid species. *Genetics*, **142**: 1357-1362.

Roques S, Duchesne P and Bernatchez L (1999). Potential of microsatellites for individual assignment: the North Atlantic redfish (genus *Sebastes*) species complex as a case study. *Molecular Ecology*, **8**: 1703-1717

Rosenberg NA (2004). DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**: 137-138.

Rottenberg H (2007). Exceptional longevity in songbirds is associated with high rates of evolution of cytochrome b, suggesting selection for reduced generation of free radicals. *Journal of Experimental Biology*, **210**: 2170-2180.

Rottenstreich S, Miller JR and Hamilton MB (2007). Steady state of homozygosity and G_{ST} for the island model. *Theoretical Population Biology*, **72**(2): 231-244.

Rousset F (2008). GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, **8**(1): 103-106.

Rousset F & Raymond M (1997). Statistical analyses of population genetic data: new tools, old concepts. *Trends in Ecology and Evolution*, **12**: 313-317.

Rozas J (2009). DNA sequence polymorphism analysis using DNAsp. In: *Bioinformatics for DNA sequence analysis; Methods* (ed – Posada D). Molecular Biology Series vol. 536. Human Press, NJ, USA.

Rozen S & Skaletsky H (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods in Molecular Biology*, **132**: 365-386.

Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V and Wallace DC (2004). Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science*, **303**: 223-226.

Russon IJ, Kemp PS and Lucas MC (2011). Gauging weirs impede the upstream migration of adult river lamprey *Lampetra fluviatilis*. *Fisheries Management and Ecology*, **18**: 201-210.

Ruxton GD (2006). The unequal variance t-test is an underused alternative to Student's t-test and the Mann-Whitney U test. *Behavioural Ecology*, **17**(4): 688-690.

Ryman N (1991). Conservation genetics considerations in fishery management. *Journal of Fish Biology*, **39** (Supplement A): 211-224.

Ryman N & Leimar O (2009). G_{ST} is still a useful measure of genetic differentiation – a comment on Jost's D. *Molecular Ecology*, **18**(10): 2084-2087.

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA and Arnheim N (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analyses for diagnosis of sickle cell anaemia. *Science*, **230**: 1350-1354.

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB and Ehrlich HA (1988). Primer - directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science,* **239**: 487-491.

Salducci MD, Martin J-F, Pech N, Chappaz R, Costedoat C and Gilles A (2004). Deciphering the evolutionary biology of freshwater fish using multiple approaches-insights for the biological conservation of the vairone (*Leuciscus souffia souffia*). *Conservation Genetics*, **5**: 63-77.

Salgueiro P, Carvalho G, Collares-Peireira MJ and Coelho MM (2003). Microsatellite analysis of genetic population structure of the endangered cyprinid *Anaecypris hispanica* in Portugal: implications for conservation. *Biological Conservation*, **109**: 47-56.

Salzburger W, Branstätter A, Gilles A, Parson W, Hempel M, Sturmbauer M and Meyer A (2003). Phylogeography of the vairone (*Leuciscus souffia*, Risso 1826). *Molecular Ecology*, **12**: 2371-2386.

Salzburger W, Ewing GB and von Haeseler (2011). The performance of phylogenetic algorithms in estimating haplotype genealogies with migration. *Molecular Ecology*, **20**: 1952-1963.

Sanger F, Nicklen S and Chase AR (1977). DNA sequencing with chain terminating inhibitors. *Proceedings of the National Academy of Sciences USA*., **74**(12): 5463-5468.

Santiago E & Caballero A (2005). Variation after a selective sweep in a subdivided population. *Genetics*, **169**: 475-483.

Schlötterer C (2000). Evolutionary dynamics of microsatellite DNA. *Chromosoma*, **109**: 365-371.

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

Schlötterer C, Amos B and Tautz D (1991). Conservation of polymorphic simple sequence loci in Cetaceans. *Nature*, **354**: 63-65.

Schneider S & Excoffier L (1999). Estimation of demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics*, **152**(3): 1079-1089.

Schulz-Mirbach T & Reichenbacher B (2006). Reconstruction of Oligocene and Neogene freshwater fish faunas – an actualistic study on cypriniform otoliths. *Acta Palaeontological Polonica*, **51**: 283-304.

Schwartz MK & McKelvey KS (2009). Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. *Conservation Genetics*, **10**: 441-452.

Sedell JR, Reeves GH, Hauer FR, Stanford JA and Hawkins CP (1990). Role of refugia in recovery from disturbances: Modern fragmented and disconnected river systems. *Environmental Management*, **14**(5): 711-724.

Šedivá A, Janko K, Šlechtova V, Kotlík P, Simonović P, Delic A and Vassilev M (2008). Around or across the Carpathians: colonization model of the Danube basin inferred from genetic diversification of stone loach (*Barbatula barbatula*) populations. *Molecular Ecology*, **17**(5): 1277-1292.

Selkoe KA & Toonen RJ (2006). Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters*, **9**: 615-629.

Serjeantson D & Woolgar C (2006). Fish consumption in Medieval England. In: *Food in Medieval England: Diet and Nutrition* (eds – Woolgar C, Serjeantson D and Waldron T). Oxford University Press, Oxford.

Sharma H, Singh A, Sharma C, Jain SK and Singh N (2005). Mutations in the mitochondrial DNA D-Loop are frequent in cervical cancer. *Cancer Cell International*, **5**: 34-40.

Sharma S & Hughes JM (2011). Genetic structure and phylogeography of two freshwater fishes, *Rhadinocentrus ornatus* and *Hypseleotris compressa*, in southern Queensland, Australia, inferred from allozymes and mitochondrial DNA. *Journal of Fish Biology*, **78**(1): 57-77.

Sheail J (1988). River regulation in the UK: an historical perspective. *River Research and Applications*, **2**(3): 221-232.

Shedlock AM, Parker JD, Crispin DA, Pietsch TW and Burmer GC (1992). Evolution of the salmonid mitochondrial control region. *Molecular Biology and Evolution*, **10**: 1259-1272.

Shpak M (2005). Evolution of variance in offspring number: the effects of population size and migration. *Theory in Biosciences*, **124**(1): 65-85.

Sjödin P, Kaj I, Krone S, Lascoux M and Nordborg M (2005). On the meaning and existence of an effective population size. *Genetics*, **169**: 1061-1070.

Slate J & Gemmell NJ (2004). Eve 'n' Steve: recombination of human mitochondrial DNA. *Trends in Ecology and Evolution*, **19**: 561-563.

Slatkin M (1977). Gene flow and genetic drift in a species subject to local extinctions. *Theoretical Population Biology*, **12**: 253-262.

Slatkin M (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**: 457 – 462.

Slatkin M & Hudson RR (1991). Pairwise comparisons of mitochondrial DNA sequences in stale and exponentially growing populations. *Genetics*, **129**: 555-562.

Smith KG & Darwell WRT (2006). *The Status and Distribution of Freshwater Fish in the Mediterranean Basin*. IUCN, Gland, Switzerland and Cambridge, UK.

Smith MW, Smith MH & RK Chesser (1983). Biochemical genetics of mosquitofish. Environmental correlates and temporal and spatial heterogeneity of allele frequencies within a river drainage. *Copeia*, **1**: 182-193.

Smith PJ, Jamieson A and Birley AJ (1990). Electrophoretic studies and stock concept in marine teleosts. *ICES Journal of Marine Science*, **47**: 231-245.

Smouse P & Peakall R (1999). Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity*, **82**(5): 561-573.

Sokal RR & Rohlf FJ (1995). *Biometry: The Principles and Practice of Statistics in Biological Research*, 3rd Edition. WH Freeman & Co, New York.

Soulé ME (1972). Phenetics of natural populations. III. Variation in insular populations of a lizard. *American Naturalist*, **106**: 429-446.

Soulé ME (1976). Allozyme variation, its determinants in space and time. *Molecular Evolution* (ed – Ayala FJ). Sinauer, Sunderland, Massachusetts, USA.

Sousa V, Penha F, Collares-Pereira MJ, Chikhi L and Coelho MM (2008). Genetic structure and signature of population decrease in the critically endangered freshwater cyprinid *Chondrostoma lusitanicum. Conservation Genetics*, **9**: 791-805.

Spiers DC & Gurney WSC (2001). Population persistence in rivers and estuaries. *Ecology* **82**(5): 1219-1237.

Stefani F, Galli P, Crosa G, Zaccara S and Calamari D (2004). Alpine and Appenine barriers determining the differentiation of the rudd (*Scardinius erythrophthalmus* L.) in the Italian peninsula. *Ecology of Freshwater Fish*, **13**: 168-175.

Stolbunov IA & Pavlov DD (2005). Behavioural differences of various ecological groups of roach *Rutilus rutilus* and perch *Perca fluviatilis*. Journal of Ichthyology, **46** (Supplement. 2): 213-219.

Stolbunov IA & Gerasimov YV (2008). Morphological and behavioural variation in juvenile roach Rutilus rutilus (Cyprinidae, Cypriniformes) from different biotopes of the Rybinskoe Water Reservoir. *Journal of Ichthyology*, **48**(2): 177-187.

Suárez J, Bautista JM, Almodóvar A and Machordorn A (2001). Evolution of the mitochondrial control region in Palearctic brown trout (*Salmo trutta*) populations: the biogeographical role of the Iberian Peninsula. *Heredity*, **87**: 198-206.

Suchard M, Weiss R and Sinsheimer J (2001). Bayesian selection of continuous-time Markov chain evolutionary models. *Molecular Biology and Evolution*, **18**: 1001-1013.

Swatdipong A, Vasemägi A, Koskinen M, Piironen J and Primmer C (2009). Unanticipated population genetic structure of European grayling in its northern distribution: implications for conservation prioritization. *Frontiers in Zoology*, **6**: 6.

Tajima F (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**: 585-595.

Takada M, Tachihara K, Kon T, Yamamoto G, Kei'ichiro I, Miya M and Nishida M (2010). Biogeography and evolution of the *Carassius auratus*-complex in East Asia. *BMC Evolutionary Biology*, **10**: 7.

Takahashi K & Goto A (2001). Evolution of East Asian ninespine stickleback as shown by mitochondrial DNA control region sequences. *Molecular Phylogenetics and Evolution*, **21**: 135-155.

Tamura K (1992). Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C content biases. *Molecular Biology and Evolution*, **9**: 678-687.

Tamura K & Nei M (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10**: 512-526.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M and Kumar S (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance and maximum parsimony methods. *Molecular Ecology and Evolution*, **28**: 2731-2739.

Tang Q, Liu H, Mayden R and Xiong B (2006). Comparison of evolutionary rates in the mitochondrial DNA cytochrome b gene and control region and their implications for phylogeny of the Cobitoidea (Teleostei: Cypriniformes). *Molecular Phylogenetics and Evolution*, **39**: 347-357.

Taskinen J & Kortet R (2002). Dead and alive parasites: sexual ornaments signal resistance in the male fish, *Rutilus rutilus*. *Evolutionary Ecology Research*, **4**: 919-929.

Taylor EB, Pollard S and Louie D (1998). Mitochondrial DNA variation in bull trout (*Salvelinus confluentus*) from northwestern North America: implications for zoogeography and conservation. *Molecular Ecology*, **8**: 1155 – 1170.

Templeton AR (1986). Coadaptation and outbreeding depression. In: *Conservation Biology* (ed – Soulé ME). Sinauer, Sunderland.

Templeton AR (1998). Nested clade analysis of phylogeographical data: testing hypotheses about gene flow and population history. *Molecular Ecology*, **7**: 381-397.

Templeton AR (2006). *Population Genetics and Microevolutionary Theory*. Wiley & Sons Inc. New Jersey.

Templeton AR, Crandall KA and Sing CF (1992). A cladistics analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. III. Cladogram estimation. *Genetics*, **132**: 619-633.

Tessier N, Bernatchez L, Presa P and Angers B (1995). Gene diversity analysis of mitochondrial DNA, microsatellites and allozymes in landlocked Atlantic salmon (*Salmo salar* L.). *Journal of Fish Biology*, **47** (Supplement A): 156-163.

Tessier N, Bernatchez L and Wright JM (1997). Population structure and impact of supportive breeding inferred from mitochondrial and microsatellite DNA analyses in landlocked Atlantic salmon (*Salmo salar* L.). *Molecular Ecology*, **6**: 735-750.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F and Higgins DG (1997) The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **24**: 4876-4882.

Tipton ML, Gignoux-Wolfsohn S, Stonebraker P and Chernoff B (2011). Postglacial recolonization of eastern Blacknose Dace (*Rhinichthys atratulus*) (Teleostei: Cyprinidae), through the gateway of New England. *Ecology and Evolution*, 1(3): 343 – 358.

Tonteri A, Titov S, Veselov A, Zubchenko A, Koskinen M, Lesbarrères, Kaluzchin S, Bakhmet I, Lumme J and Primmer CR (2005). Phylogeography of anadromous and non-anadromous Atlantic salmon (*Salmon salar* L.) from northern Europe. *Annale Zoologici Fennici*, **42**: 1-22.

Torrente-Vilara G, Zuanon J, Leprieur F, Oberdorff T and Tedesco PA (2011). Effects of natural rapids and waterfalls on fish assemblage structure in the Madeira River (Amazon Basin). *Ecology of Freshwater Fishes*, **20**: 588-597.

Trakhtenbrot A, Nathan R, Perry G and Richardson DM (2005). The importance of long-distance dispersal in biodiversity conservation. *Diversity and Distributions*, **11**: 173-181.

Travade F, Larinier M, Boyer-Bernard S & Dartiguelongue J (1998). Performances of four fish-pass installations recently built on two rivers in southwest France. In: *Fish Migration and Fish Bypasses* (eds - Jungworth, M, Schmutz S & Weiss S). Fishing News Books, Blackwell Science Ltd., Oxford.

Triantafyllidis A, Krieg F, Cottin C, Abatzopoulos TJ, Triantaphyllidis C and Guyomard R (2002) Genetic structure and phylogeography of European catfish (*Silurus glanis*) populations. *Molecular Ecology*, **11**: 1039–1055.

Tsipas G, Tsiamis G, Vidalis K and Bourtzis K (2009). Genetic differentiation among Greek lake populations of *Carassius gibelio* and *Cyprinus carpio carpio*. *Genetica*, **136**: 491-500.

Tufto J, Engen S and Hindar K (1996). Inferring patterns of migration from gene frequencies under equilibrium conditions. *Genetics*, **144**: 1911-1921.

Turner TF, Dowling TE, Broughton RE and Gold JR (2004). Variable microsatellite markers amplify across divergent lineages of cyprinid fishes (subfamily Leusicinae). *Conservation Genetics*, **5**: 279-281.

Urban D & Keitt T (2001). Landscape connectivity: a graph-theoretic perspective. *Ecology*, **82**(5): 1205-1218.

Utter FM (1991). Biochemical genetics and fisheries management: an historical perspective. *Journal of Fish Biology,* **39** (Supplement A): 1-20.

Utter FM (1994). Perspectives of molecular genetics and fisheries into the 21st century. *Reviews in Fish Biology and Fisheries*, **4**: 374-378.

Valdés AM, Slatkin M and Freimer NB (1993). Allele frequencies at microsatellite loci: the stepwise mutation model revisited. *Genetics*, **133**: 737-749.

Väli U, Einarsson A, Waits L, and Ellegren H (2008). To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? *Molecular Ecology*, **17**: 3808-3817.

Vanin ER (1985). Processed pseudogenes: characteristics and evolution. *Annual Review of Genetics*, **19**: 253-272.

Verspoor E (1997). Genetic diversity among Atlantic salmon (*Salmo salar* L.) populations. *International Journal of Marine Science*, **54**:965-973.

Vigilant L, Stoneking H, Harpending H, Hawkes K and Wilson AC (1991). African populations and the evolution of human mitochondrial DNA. *Science*, **253**: 1503-1507.

Volckaert FAM, Hänfling B, Hellemans B and Carvalho GR (2002). Timing of the population dynamics of bullhead (*Cottus gobio*) (Teleostei: Cottidae) during the Pleistocene. *Journal of Evolutionary Biology*, **15**: 930-944.

Vøllestad LA & L'Abée-Lund JH (1987). Reproductive biology of stream spawning Roach, *Rutilus rutilus*. *Environmental Biology of Fishes*, **18**: 219-227.

Vøllestad LA & L'Abée-Lund JH (1990). Geographic variation in life-history strategy of female roach, *Rutilus rutilus* (L.). *Journal of Fish Biology*, **37**(6): 853-864.

Vrijenhoek RC (1979). Genetics of a sexually reproducing fish in a highly fluctuating environment. *The American Naturalist*, **113**(1): 17-29.

Vyskočilova M, Simkova A & Martin JF (2007). Isolation and characterization of microsatellites in *Leuciscus cephalus* (Cypriniformes, Cyprinidae) and cross-species amplification within the family Cyprinidae. *Molecular Ecology Notes*, **7**: 1150–1154.

Wade MJ (2006). Natural Selection. In: *Evolutionary Genetics: Concepts and Case Studies* (eds – Fox CW and Wolf JB). Oxford University Press, Oxford.

Wade MJ & McCauley DE (1988). Extinction and recolonization: Their effects on the genetic differentiation of local populations. *Evolution*, **42**: 995-1005.

Wagner AR (1992). *Untersuchungen zur Verteilungsdynamik pelagischer Fische in funf Seen unterschiedlicher trophischerStruktur, unter besonderer Berucksichtigung der Populationsbi- ologie von Flußbarsch (Perca fluviatilis L., Pisces, Percidae) und Rotauge (Rutilus rutilus L., Pisces, Cyprinidae*). Dissertation, Johannes-Guttenberg Universita¨t Mainz.

Wakeley J (2004). Metapopulations and coalescent theory. In: *Ecology, Genetics, and Evolution of Metapopulations* (eds - Hanski I & Gaggiotti O). Elsevier, Oxford.

Wakeley J (2008). *Coalescent Theory: An Introduction*. Roberts and Company Publishers, Colorado, USA.

Wallace DC (1992). Diseases of the mitochondrial DNA. *Annual Review of Biochemistry*, **61**: 1175-1212.

Wallace DC, Brown MD and Lott MT (1994). Mitochondrial DNA sequence variation in human evolution and disease. *Gene*, **91**(19): 8739-8746.

Walsh JB (1992). Intracellular selection, conversion bias, and expected substitution rate of organelle genes. *Genetics*, **130**(4): 939-946.

Wang J (2005). Estimation of effective population sizes from data on genetic markers. *Philosophical Transactions of the Royal Society B*, **360**: 1395-1409.

Wang J & Whitlock M.C (2003) Estimating effective population size and migration rates from genetic samples over space and time. *Genetics* **163**: 429-446.

Wang Z, Zhou J, Ye Y, Wei Q, and Wu Q (2006). Genetic structure and low-genetic diversity suggesting the necessity for conservation of the Chinese longsnout catfish, *Leiocassis longirostris* (Pisces: Bagriidae). *Environmental Biology of Fishes*, **75**: 455-463.

Waples RS (1998). Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *Journal of Heredity*, **89**(5): 438-450.

Waples RS (2006). A bias correction for estimates of effective population size based on linkage disequilibrum at unlinked gene loci. *Conservation Genetics*, **7**: 167-184.

Waples RS & Gaggiotti O (2006). What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their connectivity. *Molecular Ecology*, **15**(6): 1419-1439.

Waples RS & Do C (2008). LDNE: a program for estimating effective population size from data on linkage disequilibrium. *Molecular Ecology Resources*, **8**: 753-756.

Ward RD (2000). Genetics in fisheries management. *Hydrobiologica,* **420**: 191-201.

Ward RD, Woodwark M and Skibinski DOF (1994). A comparison of genetic diversity levels in marine, freshwater and anadromous fish. *Journal of Fish Biology,* **44**: 213-232.

Warren ML & Pardew MG (1998). Road crossings as barriers to small-stream fish movement. *Transactions of the American Fisheries Society*, **127**: 637-644.

Wayne ML & Miyamoto MM (2006). Genetic variation. In: *Evolutionary Genetics: Concepts and Case Studies*. Oxford University Press, Oxford.

Weber JL (1990). Informativeness of human $(dC\text{-}dA)_n(dG\text{-}dT)_n$ polymorphisms. *Genomics*, **7**: 524-530.

Wedekind C (1996). Lek-like spawning behaviour and different female mate preferences in roach (*Rutilus rutilus*). *Behaviour*, **133**: 681-695.

Weir BS (1996). *Genetic Data Analysis II*. Sinauer, Sunderland.

Weir BS and Cockerham CC (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**: 1358-1370.

Weiss S, Antunes A, Schlötterer C and 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**: 691-698.

Weiss S, Persat H, Eppe R, Schlötterer C and Uiblein F (2002). Complex patterns of colonization and refugia revealed for European grayling *Thymallus thymallus*, based on complete sequencing of the mitochondrial DNA control region. *Molecular Ecology*, **11**: 1393-1407.

Weithman AS (1999) Socioeconomic benefits of fisheries. In: *Inland Fisheries Management in North America*, 2nd edition (eds - Kohler CC and Hubert WA). American Fisheries Society. Bethesda, MD.

Welcomme RL (2001). *Inland Fisheries: Ecology and Management*. Blackwell Science Ltd., Oxford.

Went AEJ (1950). Notes on the introduction of some freshwater fish into Ireland. *Journal of the Department of Agriculture, Dublin*, **47**: 3-8.

Wheeler A (1976). On the population of roach (*Rutilus rutilus*), rudd (*Scardinius erythropthalmus*) and their hybrid in the Esthwaite Water, with notes on the distinction between them. *Journal of Fish Biology*, **9**: 391-400.

Wheeler A & Easton K (1977). Hybrids of chub and roach (*Leuciscus cephalus* and *Rutilus rutilus*) in English Rivers. *Journal of Fish Biology*, **12**(2): 167-171.

White DJ, Wolff JN, Pierson M and Gemmell NJ (2008). Revealing the hidden complexities of mtDNA inheritance. *Molecular Ecology*, **17**: 4925-4942.

White R, Iredale R and Armstrong G (2006). Fish passes at flow measurement structures. *Proceedings ICE Water Management*, **159**: Issue WM3.

Whitehead A (2009). Comparative mitochondrial genomics within and among species of killifish. *BMC Evolutionary Biology*, **9**: 11.

Whitlock MC (1992). Nonequilibrium population structure in forked fungus beetles: extinction, colonisation and the genetic variance among populations. *American Naturalist*, **139**: 952-970.

Whitlock MC & McCauley DE (1999). Indirect measures of gene flow: F_{ST} does not equal 1/(4Nm+1). *Heredity*, **82**: 117-125.

Więsk K & Załakowski W (2000). Growth of roach *Rutilus rutilus* (L.) in the River Odra estuary. *Acta Ichthyologica et Piscatoria*, **36**(1): 4-7.

Wilklund T, Lounasheimo L, Lom J, and Bylund G (1996). Gonadal impairment in roach *Rutilus rutilus* from Finnish coastal areas of the northern Baltic Sea. *Diseases of Aquatic Organisms*, **26**: 163-171.

Wilson GA & Rannala B (2003). Bayesian inference of migration rates using mutilocus genotypes. *Genetics*, **163**: 1177-1191.

Winnepennickx B, Backeljau T and De Wachter R (1993). Extraction of high molecular weight DNA from molluscs. *Trends in Genetics*, **9**: 407.

Winter HV & van Densen WLT (2001). Assessing the opportunities for upstream migration of non-salmonid fishes in the weir-regulated River Vecht. *Fisheries Management and Ecology*, **8**: 513-532.

Wolff JN, Gandre S, Kalinin A Gemmell NJ (2008). Delimiting the frequency of paternal leakage of mitochondrial DNA in Chinook salmon. *Genetics*, **179**: 1029-1032.

Wolter C (1998). Populationsgenetische Untersuchung von Blei *Abramis brama*, Güster *Abramis bjoerkna*, Plötze *Butilus rutilus* und Rotfeder *Scardinius erythrophthalmus* aus Gewässern des nordostdeutschen Tieflandes. PhD Theses, Humboldt-University Berlin.

Wolter C (1999). Comparison on intraspecific variability in four common cyprnids, *Abramis brama*, *Abramis bjoerkna*, *Rutilus rutilus* and *Scardinius erythrophthalmus*, within and between lowland river systems. *Hydrobiologia*, **394**: 163-177.

Wolter C (2010). Functional vs scenic restoration – challenges to improve fish and fisheries in urban waters. *Fisheries Management and Ecology*, **17**: 176-185.

Wolter C, Kirschbaum F and Ludwig A (2003). Sub-population structure of common fish species in the Elbe River estimated from DNA analysis. *Journal of Applied Ichthyology*, **19**: 278-283.

Woolley S, Johnson J, Smith MJ, Crandall KJ and McClellan DA (2003). TreeSAAP: Selection on amino acid properties using phylogenetic trees. *Bioinformatics*, **19**: 671-672.

Workman PL (1969). The analysis of simple genetic polymorphism. *Human Biology*, **41**: 97-114.

Wortley J (1995) Recreational Fisheries. In: *Review of Inland Fisheries and Aquaculture in the IEFAC Area by Sub-Region and Sub-Sector* (ed –O'Grady K). FAO Fisheries Report **509**(1): 60-73.

Wright S (1931). Evolution in Mendelian populations. *Genetics*, **16**: 97-159.

Wright S (1938a). The distribution of gene frequencies under irreversible mutation. *Proceedings of the National Academy of Sciences USA*, **24**: 253-259.

Wright S (1938b). Size of population and breeding structure in relation to evolution. *Science*, **87**: 430-431.

Wright S (1940). The statistical consequences of Mendelian heredity in relation to speciation. In: *The New Systematics* (ed - Huxley J). Oxford University Press, Oxford.

Wright S (1943). Isolation by distance. *Genetics*, **28**: 139-156.

Wright S (1946). Isolation by distance under diverse systems of mating. *Genetics*, **31**: 39-59.

Wright S (1951). The genetical structure of populations. *Annals of Eugenics*, **15**: 323-354.

Wright S (1962). *Oral History Memoir*. Columbia University, Oral History Research Office, New York, 1962.

Wright S (1965). The interpretation of population structure by *F*-statistics with special regard to systems of mating. *Evolution*, **19**: 395-420.

Yan J, Liu L, Liu S, Guo X and Liu Y (2010). Comparative analysis of mitochondrial control region in polyploidy hybids of red crucian carp (*Carassius auratus*) x blunt snout bream (*Megalobrama amblycephala*). *Fish Physiology and Biochemistry*, **36** (2): 263-272.

Youngson AF, Jordan WC and Hay DW (1994). Homing of adult Atlantic salmon (*Salmo salar* L.) to a tributary stream in a major river catchment. *Aquaculture*, **121**: 259-267.

Yue GH, David L and Orban L (2006). Mutation rate and pattern of microsatellites in common carp (*Cyprinus carpio* L.). *Genetica*, **129**(3): 329-331.

Zane L, Bargelloni L and Paternello T (2002). Strategies for microsatellite isolation: a review. *Molecular Ecology*, **11**: 1-16.

Zardoya R, Economidis PS and Doadrio I (1999). Phylogenetic relationships of Greek Cyprinidae: Molecular evidence for at least two origins of the Greek cyprinid fauna. *Molecular Phylogenetics and Evolution*, **13**: 122-131.

Zhang DX & Hewitt GM (1996). Nuclear integrations: challenges for mitochondrial DNA markers. *Trends in Ecology & Evolution*, **11**: 247-251.

Zimmerman DW (1987). Comparative power of the Student *t* test and Mann-Whitney *U* test for unequal sample sizes and variances. *Journal of Experimental Education*, **55***:* 171-174.