UNIVERSITY OF HULL

Genomic architecture of *Triops cancriformis* sex chromosomes in response to sexual system transition

A thesis submitted for the degree of MSc by Research

in the University of Hull

by

David Thomas Orr, BSc (Hons)

September 2017

Acknowledgements

Data used for this study was obtained by Thomas Mathers.

I would like to thank my supervisors Africa Gómez and Christoph Hahn for their continued support and encouragement during this masters. I would especially like to thank Africa Gómez as her infectious enthusiasm and inspiration has continually kept me motivated and encouraged.

Also thanks to the rest of the EvoHull team who have been always been helpful, supportive and provided coffee and feedback.

I would also like to thank my partner Jessie who has given an immeasurable level of support thoughout my studies.

Table of Contents

Abstract

Sexually antagonistic selection has been suggested to be a leading factor in the reduction of intralocus sexual conflict and driver of sex chromosome evolution. Although the scarcity of empirical data makes it difficult to quantify the impact sexually antagonistic selection has on sex chromosome evolution. *Triops cancriformis*, have gonochoric and androdioecious (hermaphrodites and rare males) populations giving us a unique opportunity to assess the impact of sexually antagonistic selection across different sexual systems with different pressures of sexual conflict. We aim to characterise sex chromosome structure, differentiation and gene content in *T. cancriformis* populations with different sexual systems. We used paired-end RAD sequence data of 47 different individuals from two populations (gonochoric and androdioecious), in combination with two new genomic assemblies of *T.cancriformis*. We increase the number of putative sex linked genetic markers from 52 to 79, finding 18 contigs that are sex linked in both populations to a high confidence level. We accurately characterised the sex chromosome, determining that the sex chromosomes are most likely homomorphic. We find a smaller region under recombination suppression in the androdioecious population, divergent Z chromosomes between populations and determine two different lineages of W chromosome present in the androdioecious population. We also show evidence of a possible recombination event in the androdioecious population and determine a panel of genes that could be involved in sex determination.

Introduction

The composition of sexes within a population or species determine their sexual system. Gonochorism (also known as dioecy), is characterised by the presence of only male and females in a population, is the most common sexual system found in animals, unlike in plants where hermaphroditism, individuals producing both female and male gametes is predominant (Jarne & Auld, 2006; Weeks, 2012). Gonochorism is thought to be the derived sexual system in the animalia kingdom, only 5% of species are hermaphroditic, although this rises to nearly a third of known animal species when discounting the bias from the species-rich arthropoda (Bachtrog et al., 2014; Ghiselin, 1969). Hermaphroditism is widespread in invertebrates, occurring in molluscs, reef building corals and branchiopoda but in vertebrates it is generally confined to fish (Avise and Nicholson, 2011). Hermaphroditism consists of two categories: sequential hermaphroditism, seen when an individual's sex changes during its lifetime; and simultaneous hermaphroditism, where an individual produces both female and male gametes throughout its lifetime (Beukeboom & Perrin, 2014). Obligate self-crossing (selfing) can be a consequence of simultaneous hermaphroditism and although it can increase colonisation effectiveness through reproductive assurance, it can have negative impacts on genetic variation within populations as well as effective population size and efficacy of selection (Charlesworth & Wright, 2001; Jordan & Connallon, 2014).

As gonochorism is more common, a higher frequency of transitions from gonochorism to the ancestral animal state hermaphroditism is observed, and these transitions to hermaphroditism from gonochorism tend to enable self-fertilisation (Weeks et al., 2014). Self-fertilisation enhances an individual's ability to achieve reproductive assurance, especially during changes to the environment, either through colony establishment and recolonisation of habitats or through low and fluctuating metapopulations (Pannell, 2002; Pannell & Barrett, 1998). These self-fertilising hermaphrodites that have evolved from a single sex tend to have a biased morphology towards female (Beukeboom & Perrin, 2014). This is expected due to low constraints needed to evolve the male gamete production necessary to perform self-fertilisation in comparison to the suite of secondary sexual characteristics such as egg laying and

maternal care, which males would need to obtain for self-fertilisation. Transitions between gonochorism and hermaphroditism are thought to pass through an intermediate, transitional stage consisting of hermaphrodites and a low proportion of females (gynodioecy) or hermaphrodites and a low proportion of males (androdioecy) (Charlesworth and Charlesworth, 1978). The transitional states can be maintained as a stable system if the hermaphrodite sex cannot evolve mechanisms to initiate outcrossing and males or females are maintained in populations due to their reduced inbreeding depression (Weeks et al., 2006). Due to the greater anatomical dissimilarity between males and hermaphrodites in animals, compared to females and hermaphrodites, males can fill and occupy a particular ecological niche. In plants this dissimilarity is less pronounced and the selection for separate sexes comes from inbreeding avoidance pressure. Research in plants show, as male gametes are in surplus to female gametes and males suffer from a reduction in mating opportunities with the appearance of hermaphrodites (Sakai et al., 1997). Females will not be subjected to the same reduction in mating opportunities as they can be fertilised by the male gametes produced by hermaphrodites, resulting in the preservation of females over males during the transition from gonochorism to hermaphroditism. Theoretically, this would result in androdioecy being a more frequent transitional state within animals than plants and this is supported by the scarcity of gynodioecy in comparison to androdioecy in animals (Weeks, 2012).

Sex Determination

There are numerous mechanisms which determine whether an individual develops as either a male or female. Environment factors, such as temperature, determine the sex of the offspring in many reptiles and some fish (Devlin & Nagahama, 2002; Janzen & Paukstis, 1991). This trait can be advantageous if a particular sex can benefit from certain conditions (Bull, 1983), but it can be problematic, with climate change having detrimental effects on sex ratios and effective population sizes (Mitchell & Janzen, 2010). In genetic sex determination (GSD), sex is determined through an inherited genotype. A number of GSD mechanisms occur: XY determination system - ubiquitous across all placental mammals and model species including humans and *Drosophila melanogaster*. Often mistakenly identified as the typical sex determination system, sex is determined by the presence of sex chromosomes in which females are

the homogametic sex (XX) and males are heterogametic (XY) (Bachtrog et al., 2014; Pan et al., 2015). ZW determination system - sex is also determined by sex chromosomes but in which males are the homogametic sex (ZZ) and females are the heterogametic sex (ZW), this is present in birds and Lepidoptera (Traut and Marec, 1997; Zhou et al., 2014). Some lineages determine sex through the absence of a sex chromosome, this is seen in the XO or ZO sex determining systems as well as haplodiploidy, which is ubiquitous in the hymenoptera and enables the control of sex ratios in populations, as only fertilized eggs develop into females and unfertilized eggs develop into males (Beukeboom & Perrin, 2014). Although GSD systems are strictly determined by genetic factors, they can be manipulated and overridden by non-genetic factors. These epigenetic sex determination factors include parasite manipulation, meiotic sex drivers and social cues (Bachtrog et al., 2014; Rutkowska and Badyaev, 2008; Úbeda et al., 2014). For example, Wolbachia is an intracellular bacterial parasite, which is transmitted through females. Its presence can cause feminisation, parthenogenesis, male killing and cytoplasmic incompatibility across male arthropods, overriding the GSD which increases the female sex ratio and, thus, increases the number of viable hosts available for transmission (Werren et al., 2008).

In some model species possessing GSD, such as *Caenorhabditis elegans, Drosophila melanogaster* and mammals, a monogenic master-switch sex determining system is present, in which the sex determination is controlled by a single gene (Bachtrog et al., 2014). This is not the case for all organisms with GSD, for example, in the zebrafish sex determination is polygenic and sex is determined by multiple loci in the genome (Anderson et al., 2012). Also, genes involved in monogenic master-switch sex determining systems can be very conserved throughout the taxa, (e.g. *Sry* gene is almost ubiquitous throughout placental mammals), or can be more labile, with closely related species possessing different genes (Foster & Graves, 1994; Myosho et al., 2012).

Sex Chromosome Evolution

Sex chromosomes have evolved independently numerous times in plants and animals, giving biologists the opportunity to study the varying evolutionary forces across a large range of taxa (Bachtrog et al., 2011). It is now widely believed that sex

chromosomes evolve from an autosomal pair carrying a genetic sex determining gene or region (Rice, 1996). The most widely accepted theoretical model of sex chromosome evolution predicts that the influence on sexual conflict at loci promotes differentiation of sex chromosomes (Wright et al., 2016). For instance, if, in an ancestral species, an allele arose that negatively impacted one sex but benefitted the other (sexually antagonistic allele), it would only become established if the benefit to one sex would outweigh the cost to the other sex. This is not the case for individuals in which the sexually antagonistic allele locus is in close proximity to sex determining locus on the sex chromosome, this tight linkage of the sexually antagonistic allele and sex determining allele would cause the allele to be only expressed in one sex (Arnqvist & Rowe, 2005). Through sex-biased gene transfer, accumulation of sexually antagonistic alleles can occur despite the highly deleterious impact it can have on a particular sex. When the sex determining region (SDR) contains sexually antagonistic alleles, selection favours the suppression of recombination to prevent the recombining of sexually antagonistic alleles and sex determining alleles. In some lineages, this recombination arrest can cause heteromorphic chromosome regions, first encompassing a small region on the sex specific chromosome but eventually the chromosome can become a hotspot for sexually antagonistic alleles that can eventually cover a large proportion of the chromosome, promoting the evolution of recombination suppression over almost the entire chromosome (Rice, 1987; Wright, 2016). This is what has occurred on the human Y chromosome, recombination between X and Y chromosomes has been increasingly suppressed with only 2% of the chromosome now in the pseudoautosomal region (PAR), the area of the chromosome which is subjected to recombination in males (Ross et al., 2005). Recombination suppression or arrest can occur through transposons, repetitive sequences and inversions of autosomal material, the consequences of this recombination arrest between alleles that are epistatically linked is the formation of a region of the genome that contain multiple loci that are then inherited together called ‘supergenes’. ‘Supergenes’ can trap genes within, usually fertility promoting genes which enhance canalisation of sexes (Beukeboom & Perrin, 2014). ‘Supergenes’ can occur along the sex chromosomes in a stepwise manner, allowing biologists to deduce a ‘evolutionary strata’ of when recombination arrest was initiated across many taxa (Zhou et al., 2014). The genetic architecture of sex chromosomes can give a particular sex an advantage in regards to intralocus sexual conflict. For instance, sexually

antagonistic, rare recessive alleles that benefit the heterogametic sex can easily spread as they will rarely be expressed in the homogametic sex due to phenotypic masking by the dominant allele in the homogametic sex. This theory predicts that the X chromosome in an X/Y or Z in a Z/W system may become enriched with alleles benefitting the heterogametic sex (Arnqvist & Rowe, 2005).

Differentiated sexes can lead to sexual conflict between sexes or the emergence of secondary sexual traits, the accumulation of these factors can drive changes to the genomic architecture of a species or population (Arnqvist & Rowe, 2005). For instance, a rare, recessive allele favouring the heterogametic sex will result in the allele being negatively selected for in the homogametic sex, although as the allele will not be expressed in the homogametic sex due to dominant masking. This allows the frequency to increase in the population and can select for heterozygosity at sex linked alleles. But this can also result in, for instance, a sexually antagonistic, female biased allele being present on a Z chromosome (Lee et al., 2002). The dynamics of this effect can be increased in androdioecious and gynodioecious populations, where one particular sex is rare. Recessive, hermaphrodite biased alleles are negatively selected for in males or females, but as these individuals are rare in a population, the frequency of hermaphrodite biased alleles can increase rapidly. It is thought, that through outcrossing the hermaphrodite will in fact decrease the fecundity of its offspring through subjection to a dominant male or female biased allele (Chasnov, 2010). These transitional sexual systems are thought to be instead supported by the extremely high selection pressures on mating in males and inbreeding depression avoidance, that is common among hermaphroditic populations (Weeks et al., 2000).

The classical view of sex chromosome evolution is that the absence of recombination can have deleterious effects on the non-recombining chromosome leading to its eventual elimination. This hypothesis suggests that recombination suppression exposes the chromosome to Hill-Robertson interferences - further lowering of effective population sizes and power of selection through interaction between linked genes (Hill & Robertson, 1966), Muller's ratchet - the accumulation of deleterious mutations without the ability to purge due to the lack of recombination (Charlesworth & Charlesworth, 2000), genetic hitchhiking and selective sweeps, whilst also increasing the impact of genetic drift (Beukeboom & Perrin, 2014). The harbouring of deleterious mutations and accumulation in non-coding DNA can lead to the degradation of sex

specific neo-chromosomes and result in differentiation of morphology between sex chromosomes (Beukeboom & Perrin, 2014). Non-recombining regions in W/Y chromosomes are usually composed of few genes, usually with sex-specific functions, and largely made up of repetitive and heterochromatic DNA (Charlesworth & Charlesworth, 2000; Skaletsky et al., 2003). This is evident in *Drosophila miranda* where the Y-chromosome, which is achiasmatic (recombination is non-existent throughout the chromosome), has incurred a drastic loss of diversity when compared to the X-chromosome with prominent decay of coding regions in the Y-chromosome (Bachtrog & Charlesworth, 2002; Bachtrog et al., 2008). Degradation of the non-recombining region of sex chromosomes suggest that the elimination of the sex chromosome is inevitable, Steinemann & Steinemann (2005) went so far as to predict that the Y-chromosome may be 'born to be destroyed'. Studies into some lineages do support this theory with some lineages of rodent, *Drosophila* and many groups of arthropod completely lacking the ancestral Y chromosome, resulting in an XX/XO sex determining system (Beukeboom & Perrin, 2014). Complete suppression of recombination does initiate the degradation of sex chromosomes and this has led to speculative estimations of lifespans of sex chromosomes (Aitken & Marshall Graves, 2002).

Recent discoveries have led to a modern view of sex chromosome evolution which show degradation of a chromosome is not linear with a multitude of forces in action to limit, compensate and reverse the degradation, disputing the fatal fate that was hypothesised for sex chromosomes (Bellott et al., 2014). Dosage compensation mechanisms are present in mammals and *Drosophila* which accommodate the decay through increased gene expression regulation, promoting the expression of underrepresented genes located in the decaying regions (Beukeboom & Perrin, 2014; Zhou et al., 2013). Birds show varied tempos of W chromosome decay across lineages since the initial inversion which initiated recombination suppression in the ancestor of all birds (Zhou et al., 2014), some Palaeognathae have evolved sex biased gene expression, eliminating the necessity of recombination suppression and allowing large proportions of the chromosome to remain in the PAR (Vicoso et al., 2013). Some lineages and groups, such as some tree frogs and most paleognaths, have retained homomorphic sex chromosomes despite recombination suppression (Stöck et al., 2011, Zhou et al., 2014). Homomorphic sex chromosomes can be observed in recently

diverged sex chromosomes, as is seen in wild strawberry species, *Fragaria chiloensis* (Tennessen, 2016), but can also be maintained due to low levels of recombination between sex chromosomes. European tree frogs evolved X-Y recombination suppression in males at least 5 Mya but have homomorphic sex chromosomes that undergo occasional recombination events during the development of occasional sex-reversed XY females, opposing the suggestion that sex chromosomes accumulate degradation until their eventual elimination (Stöck et al., 2011).

Overview of Notostraca

Notostraca are a globally distributed order of branchiopod crustaceans consisting of two genera, *Triops* and *Lepidurus*, well renowned for their extreme morphological conservatism and diapausing cysts (Fryer, 1988; Thiéry, 1997). Notostraca are found on every continent except Antarctica, residing in temporary pools of fresh and brackish waters, surviving long droughts or frost through their highly resistant dormant eggs laid in sediment and achieving wide dispersal ranges through vectors such as birds or mammals (Longhurst, 1955; Vanschoenwinkel et al., 2012). Their morphological conservatism is such that fossils dating back from the Upper Triassic period were assigned to a subspecies of currently extant species (Trusheim, 1938), this has since been refuted by recent reviews using molecular clock analyses (Korn et al., 2013; Mathers et al., 2013a). Despite this extreme morphological conservatism, intraspecific variability can be high which has led to difficulties in species identification. Longhurst (1955) named nine species of Notostraca in his revision of the order, but this number has risen to nearly 60 from recent studies utilizing molecular analyses (Korn et al., 2013; Mathers et al., 2013a; Meusel & Schwentner, 2016).

Sexual systems exist in various forms throughout the Notostraca: gonochorism occurs in most species and is the most likely ancestral sexual system (Mathers et al., 2013b). In gonochoric species, offspring production requires sex between both sexes and populations usually have balanced sex ratios (Korn et al., 2006). Selfing hermaphroditism occurs in some species, which possess the ability to reproduce in isolation. Hermaphrodites possess ovo-testes, and viable eggs are produced when grown in isolation (Zaffagnini & Trentini, 1980). Finally, the rare sexual system, androdioecy is present in some species, with populations being composed of large

proportions of selfing hermaphrodites in addition to few males. Outcrossing in these populations can only be achieved through reproduction with males (Weeks et al., 2006). Androdioecy has evolved a minimum of five times in Notostraca, this variability of sexual systems, not only between but also among species, makes Notostraca exceptional models to study the effects of sexual system transitions on the evolution of sex chromosomes (Mathers et al., 2013b).

The European tadpole shrimp, *Triops cancriformis*, occupies temporary freshwaters, ponds and rice fields throughout Europe. *T. cancriformis* populations exist in all forms of sexual systems found in Notostraca, with sex ratios in some populations in the Iberian Peninsula determined to be approximately 50:50, whereas populations in more northern and central Europe are found to be hermaphroditic and androdioecious (Zierold et al., 2007), this makes them particularly good model candidates for studies when conducting population and genomic investigations involving sexual system transitions.

Triops cancriformis sex determination

Sex determination in Notostraca was first proposed to involve a male recessive system through the analysis of inheritance among males and hermaphrodites of *Triops newberryi*, suggesting a ZW sex determination system (Sassaman,1991). This was also shown to be present within *Eulimnadia texana*, another branchiopod crustacean exhibiting androdioecy (Sassaman and Weeks, 1993). Sassaman (1991) also determined there are two types of hermaphrodites within the androdioecious populations, monogenic (possessing WW chromosomes) and amphigenic (ZW). Morphological chromosomal characterisation was performed through karyotype analysis, resulting in the identification of a 2n = 12 chromosomal system with no distinct morphological differences to identify sex chromosomes (Ombretta et al., 2005). The absence of understanding of sexual system mechanisms in branchiopod crustaceans along with their unique sexual system transitions makes them a very appealing class to study.

Mathers et al. (2015) utilised restriction site-associated DNA sequencing (RAD-seq) to compare genome-wide variation in 47 individuals consisting of males, females and hermaphrodites from a dioecious population from Espolla, Spain (ESP) and an

androdioecious population from Königswartha, Germany (KOE) in absence of a reference genome. Using a set of criteria consistent with expected chromosomal frequency patterns, Mathers et al. (2015) identified 52 candidate sex-linked markers and confirmed that *T. cancriformis* possess a ZW sex determination system which is composed of heterogametic (ZW) females, homogametic (ZZ) males with two types of hermaphrodites in the androdioecious populations (monogenic WW and amphigenic ZW). The study also revealed some intriguing features of the genomic architecture in *T. cancriformis*, indicating that similarities between the W chromosome in ESP and the Z and W chromosomes in the KOE and also that a transition to hermaphroditism may have produced an increase in recombination, or a decrease in recombination suppression between sex chromosomes. Differences in some sex linked regions also suggest that elements of the Z chromosome in the KOE have originated through introgression with another population, a hypothesis that is consistent with the occurrence of two mtDNA lineages in this population (Zierold et al, 2007).

RAD sequencing and genome sequencing technologies

Reduced representation sequencing, such as RAD-seq, has revolutionized the field of genomic research and has the ability to produce thousands of unbiased, polymorphic markers in parallel across tens or hundreds of individuals without incurring the same timescale and financial costs that arise with whole genome sequencing (Davey et al., 2010; Willing et al., 2011). RAD-seq has proved a very useful method due to its ease of reproducibility and suitability for non-model organisms for which no other genomic resources exist (Baird et al., 2008; Davey et al., 2010). There are many ways of performing RAD sequencing but the result is the formation of short sequences of loci found at intervals throughout the genome called ‘RAD tags’. Baird et al. (2008) method of RAD-seq first involves digesting genomic DNA using a restriction enzyme and ligating a P1 adaptor (containing a primer site, amplification site and barcode) onto the resulting ‘sticky-ends’. The resulting fragments, for a number of individuals, are then pooled and sheared to create smaller fragments, these smaller fragments are then ligated to a P2 adaptor (which has a divergent structure) and are then amplified using the P1 and P2 primers. The divergent P2 adaptor structure ensures that the amplification will only occur for fragments containing P1 adaptors and so each marker can be identified using its barcode. High throughput, parallel Illumina sequencing is

then performed on barcoded small fragments to create a 'RAD Seq library' in which the sequenced fragments (RAD tags) are stored (Baird et al., 2008; Davey & Blaxter, 2010).

Since its development RAD-seq has been applied to investigate number of biological problems, such as phylogeography and selection, through detection of single nucleotide polymorphisms (SNPs) in RAD markers (Emerson et al., 2010; Hohenlohe et al., 2011). Downstream analyses have been implemented through a number of bioinformatic software tools that have been created specifically for RAD-seq data analysis (Baxter et al., 2011; Catchen et al., 2013; Chong et al., 2012). RAD-seq data analysis can be used to infer population statistic and sex linked markers consistently *de novo*, i.e. in the absence of a reference genome, as performed by Mathers et al. 2015. However, when analysing RAD-seq data an appropriate reference genome, e.g. from a closely related species, should be used to improve accuracy of results in variant calling and population statistics (Shafer et al., 2016). The use of a reference genome allows the identification of loci linked to RAD tags of interest across the genome and the ability to utilise the 2nd end reads to call SNPs, which could likely increase the number of called SNPs significantly.

Aims and objectives

Here, we investigate the genomic effects on sex chromosomes occurring during the transition of sexual systems in *T. cancriformis*, combining the RAD-seq data produced by Mathers et al. (2015) with two new reference genome assemblies, one from a male *T. cancriformis* from the Spanish population (possessing a ZZ chromosomal arrangement) and another from a hermaphroditic individual from Scotland (possessing a WW chromosomal arrangement). The novelty of using a reference genome as a template will allow greater accuracy in SNP calling and the ability to determine sex linked contigs across the two sexual systems. Through further investigation of these contigs we will:

1. Determine the size of the sex linked region of the genome in the different populations. We will also look at the non recombining regions across the

populations and compare the size of these regions in two different sexual systems.

2. Mathers et al., 2015 and Zierold et al., 2007 suggested different origins of Z chromosomes present, we will compare haplotypes across populations to infer lineages.
3. The use of reference genomes has enabled us to identify genes present in the sex chromosome regions, including genes putatively involved in sex determination that could be vital in the transition between gonochorism and hermaphroditism or could be putative targets of sexual antagonistic selection.

Methods

RAD Library Sequence Data

We used published RAD sequence data downloaded from the Sequence Read Archive through the accession number PRJEB7851, detailed in Mathers et al. (2015). This sequence data, consisting of four paired-end RAD libraries, was produced by the NERC/NBAF facility at the University of Edinburgh utilising the protocol of Baird et al. (2008) with some modifications (Ogden et al. 2013). This sequenced data was composed of 47 mature, sexed *T. cancriformis* individuals from two populations: 14 males and 12 females were obtained from a dioecious population in Espolla, Spain (ESP) along with 15 hermaphrodites and 6 males obtained from an androdioecious population in pond 12 in Königswartha, Germany (Zierold et al., 2007, 2009) (Appendix, Table 1). In brief, the RAD protocol involved digesting the genomic DNA using an *Sbf1* enzyme and ligating with a P1 RAD adaptor barcode, genomic DNA from each individual was sheared and size-selected, between 300 and 700 bp, before undergoing the ligation of P2 RAD adaptors. The four libraries were amplified through PCR before being sequenced in separate flow cells on an Illumina HiSeq 2000 platform with 100 bp paired-end chemistry, two sequencing runs were performed.

Table 1. Summary of statistics of the two genome assemblies utilised for the mapping of RAD sequence data.

Genome file	**Scottish_DISCOVAR.min500.fasta**	**T_ma_SPADES_hybr_PB-5000_min500.fasta**
Population	Caerlaverock, Scotland	Espolla, Spain
Sex	Hermaphrodite	Male
Sequencer	Illumina MiSeq	PacBio RS II/ Illumina Miseq
Assembly software	DISCOVAR	SPADES
Estimated genome size (Mb)	114.99	135.28
Number of scaffolds	10022	15265
Mean scaffold size (bp)	11473	8862
N50 scaffold length (bp)	56040	41668
Longest scaffold (bp)	479734	692116

Genome assemblies

Two reference genomes were utilised for the experiment (Table 1). Assemblies were carried out by C. Hahn in the EvoHull Genomics Lab, University of Hull. The male library was sequenced on a single Illumina MiSeq sequencer run with v3 chemistry (2x300 bp paired end read lengths), library preparation and Illumina MiSeq sequencing was carried out at the Nucleus facility at the University of Leicester. To improve the assembly long read sequence data was obtained, sequenced on a PacBio RS II machine (Pacific Biosciences, CA), at the Centre for Genomic Research at the University of Liverpool. PacBio reads were corrected with unitigs produced by the Celera assembler from the MiSeq data using the ECTools correction pipeline. PacBio corrected reads and Illumina MiSeq data was assembled as a hybrid assembly using SPAdes (Antipov et al., 2016; Nurk et al., 2013). PacBio reads less than 5000 bp were used as single end reads whereas longer reads were primarily used for scaffolding using the --pacbio function. Another genome assembly was generated from a hermaphrodite individual from Caerlaverock (Scotland) using a single Illumina MiSeq sequencer run with v3 chemistry (2x250 bp paired-end read lengths). Library preparation and sequencing was carried out at the EvoHull Genomics Lab, University of Hull. The hermaphrodite genome assembled using DISCOVAR de novo (Love et al., 2016). Only contigs with a minimum length of 500 bp were retained for the final assemblies.

Demultiplexing, Quality control and Mapping

Stacks v1.44 program *process_radtags* (Catchen et al., 2013) was used to perform initial checks on the RAD raw sequence data before demultiplexing and outputting the demultiplexed files in the fastq format. *Process_radtags* was also used to discard any reads that contained an uncalled base or did not contain the expected *Sbf1* cut site or specified corresponding barcodes, with an allowance of a mismatch of up to one base to account for sequencing error (Baxter et al., 2011). Sequence quality was reassessed, raw reads were only retained if they possessed a phred quality score of at least 20 before being truncated to 70 bp and saved as Fastq files.

Due to the random shearing step, paired-end reads from different DNA molecules will likely be different lengths and so multiple copies of the exact same length are most

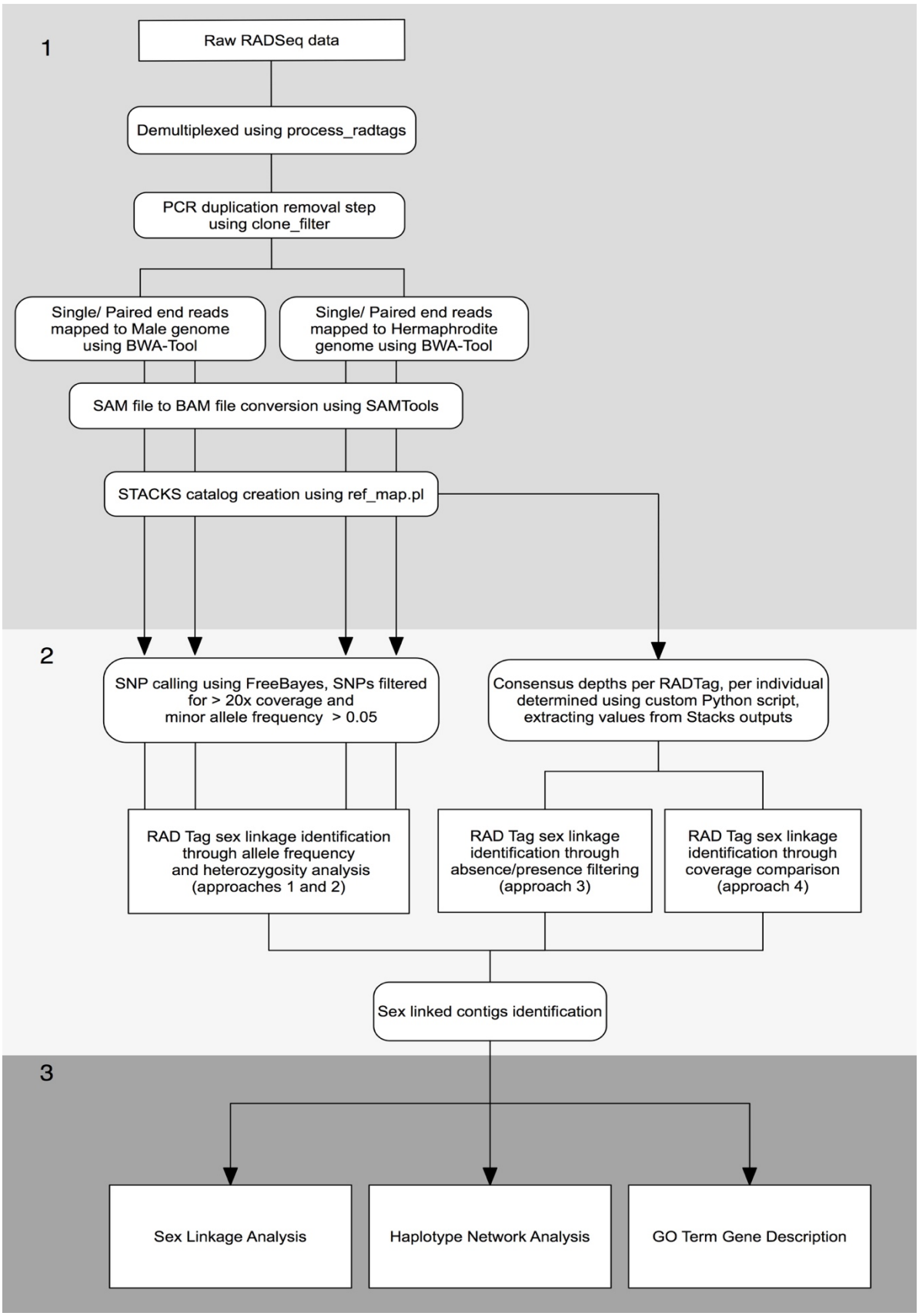


Figure 1. The flow chart used to process the RAD data.

likely to be PCR duplicates. PCR duplicates created during the amplification process of the library preparation can have an effect on the downstream analysis, resulting in inaccuracies in allele analysis and genotyping errors. To compensate for biased amplification of alleles during the PCR of genomic DNA (Andrews et al., 2016) the *clone_filter* program in Stacks v1.44 (Catchen et al., 2013) was used. In *clone_filter* identically matching single and paired-end reads are considered clones and all but one representative of the set is retained for downstream analysis. GC bias can also occur during the PCR step as sequences with different GC content are unequally amplified during PCR, the use of *clone_filter* will also eliminate the effects of these bias (Andrews et al., 2016).

The filtered reads were separately mapped to both reference genomes using a custom script (Supplementary, Figure 9.1). BWA version 0.7.15 (Li and Durbin, 2009) was first used to align the filtered reads to the reference genome using the *aln* and *samse* functions or the *aln* and *sampe* functions for single-end reads or paired-end reads respectively, mismatches of up to six base pairs were allowed and the aligned reads were outputted as SAM files. During this process a perl script *split_sam.pl* (Supplementary, Figure 9.2) was used to record the number of unique hits, multiple hits or no hits per individual. SAMTools (Li et al., 2009) was then used to convert the SAM files to ordered BAM files for downstream analysis required by the SNP calling program.

RAD Tag Catalog Creation

RAD tags were identified separately for each reference genome. The filtered RAD data from the first ends only was analysed using Stacks v1.44 (Catchen et al., 2013) and run through the *ref_map.pl* to create a catalog of RAD tags present in at least one individual (hereafter this is referred to as 'the catalog'). *Ref_map.pl* pipeline runs through the different components *pstacks, cstacks* and s*stacks* of the Stacks pipeline. BAM files were analysed through *pstacks*, which extracts 'stacks' of loci, with a minimum depth of 20 identical reads to report a 'stack'. *Cstacks* was then used to merge alleles and form a catalog of consensus loci across all individuals, allowing up to six mismatches between alleles. *Sstacks* then searched the catalog, created in

cstacks, to report the presence of each RAD tag across individuals and identify the corresponding reads.

SNP Calling

Using the outputs of the *ref_map.pl* and a population map, *populations* program was then run to output the single nucleotide polymorphisms (SNPs) and population statistics, such as F statistics. Only SNPs present in 75% of individuals, with a coverage of ≥ 20x and a minor allele frequency >0.05 were used. This procedure was performed separately for each population data, Espolla and Königswartha, and repeated for each reference genome.

Stacks will not merge indels and will identify them as separate alleles/ RAD tags, so FreeBayes (Garrison and Marth, 2012) was also used to call SNPs from the paired end RAD data, utilising both 1st and 2nd end reads. FreeBayes uses bayesian inference on the short read alignments (BAM files) to detect genetic variants within the alignments. FreeBayes also takes into account that identical sequences may have multiple possible alignments and so can identify indels among alleles. To ensure the SNPs called by FreeBayes were still adhering to the original filters set in Stacks, only read pairs containing 1st end reads that contributed to the loci in the Stacks catalog (see above) were used in the SNP calling and in further downstreams analyses. Using a custom Python script (Supplementary, Figure 9.3), a whitelist of loci from the catalog created by Stacks was parsed; the corresponding reads from these tags were then extracted from the original raw reads; these reads were concatenated into new FASTA files and then remapped to the reference genome using BWA and SAMTools to create BAM files, required by FreeBayes, before undergoing SNP calling using the *--no-complex* and *--use-best-n-alleles 2* filters (Figure 1). These filters ensured that adjacent SNPs were not merged into multi-nucleotide polymorphisms (MNPs) and that only the top 2 alleles per SNP are evaluated. After the VCF (variant call format) file of SNPs was produced, the files were filtered using VCFTools (Danecek et al., 2011), to ensure that only SNPs present in 75% of individuals, with a coverage of ≥ 20x and a minor allele frequency >0.05 were retained.

Depth Coverage

The depth of coverage observed between individuals is highly variable (Appendix, Figure 1), which is often the case when using RAD-seq data. To perform accurate

comparison of RAD tag read depths between individuals it is necessary to normalise the data (Böhne et al. 2016). Firstly, a median depth value per individual for putatively autosomal loci was derived using only the RAD tags which were universally present among individuals, had a minimum depth of coverage of 20x per individual and a mean coverage of less than 500x across individuals (Beissinger et al., 2013). This ensures sex and population specific loci are not utilised in the median coverage as well as accounting for repetitive DNA. These putatively diploid autosomal median values for each individual are used to normalise the individual's depth for each loci relative to the mean values so that coverage of 1 would be expected.

Coverage data for each RAD tag from the catalog per individual was compiled using a custom Python script (Supplementary, Figure 9.4), extracting the depth values for each RAD tag per individual from the individual's corresponding "***.matches.tsv.gz" file outputted by Stacks. The depths were then exported to an Excel spreadsheet to be filtered and normalised as described above.

RAD tag Sex Linkage Identification

We used a four-step approach to identify sex linked markers and SNPs within populations, which is a modification of the three-step approach (approaches 1-3), devised by Brelsford et al. (2016), including an additionally devised approach 4 to suit a ZW sexual system and the presence of monogenic hermaphroditic individuals. These approaches will allow us to conservatively identify putative RAD tags that are consistent with expected sex linked patterns (Table 2).

1. *Allele frequency approach:*

Given that in the ZW sexual system males are the homogametic sex, it is expected that Z specific alleles would have two copies in males and one copy in females. We compared SNP allele frequencies across sexes using VCFTools (Danecek et al., 2011). The VCF file was filtered to identify SNPs with allele frequencies ≥ 0.95 in one sex and a frequency difference of > 0.4 between the allele frequency in both sexes. Following Brelsford et al. (2016), we allowed a small deviation from the theoretical 100% frequency and 50% difference between allele frequencies in both sexes to account for sequencing errors, rare recombination events or polymorphism within sex linked RAD tags.

2. *Heterozygosity approach:*

Because males are the homogametic sex, sex-linked loci in males should possess an excess of homozygosity in comparison to females. The VCF file was filtered for SNPs where at least ⅔ of individuals from one sex are heterozygous and all individuals from the other sex are homozygous. Although this approach will return very similar results to approach 1, it will avoid false positive identification. For example, if at a locus, half of females are homozygous for one allele, half of females are homozygous for an alternate allele and all males are homozygous for one allele, this will equate to an allele frequency ≥ 0.95 in males and a frequency difference of > 0.4 between the sexes despite not adhering to sex linkage expectations (Table 2). Only loci that satisfy both approaches, 1 and 2, were considered as putatively sex linked.

3. *Presence/Absence approach - Z/W specific RAD tags:*

Mutations affecting the restriction site can lead to dropouts of particular alleles from a RAD dataset (null alleles). In addition, long periods of recombination suppression can lead to strong divergence between gametologs and increase in frequency of polymorphisms, and thus sex linked RAD tags are particularly prone to allele dropouts (Andrews et al., 2016). The catalog of RAD tags was filtered to identify RAD tags that are present in one sex but absent from the other, or present in amphigenic hermaphrodites and males while absent in monogenic hermaphrodites (see below) for Z specific RAD tags in Königswartha. Afterwards, the normalised depths of coverage of these putative sex specific RAD tags were checked, with differences expected between sexes in populations, for example a W specific RAD tag in a female or amphigenic hermaphrodite individual is expected to have 0.5x autosomal coverage (Table 2).

4. *Coverage approach - Z specific RAD tags:*

Z specific RAD tags were expected to be present in both sexes in the dioecious Espolla population, and cannot be identified by approach 3, but they should differ in average normalised coverage. To identify Z specific RAD tags, we plotted the normalised median depth of RAD tag read coverage of males against that of females. RAD tags exhibiting the 1x male/ 0.5x female expected pattern were further analysed to check that the pattern held consistently at an individual level, comparing normalised depths per sex per RAD tag per individual. Filtering was applied for RAD tags with

similar read depth among individuals of different sexes (e.g. all males had similar 1x coverage and all females similar 0.5x coverage) and conservative differences in depth coverage of >0.35x median coverage between sexes (Mathers et al., 2015).

Table 2. Table of expectations consistent with sex linked SNPs/ RAD tags. * Indicates that this data was used in the sex linked SNP approaches.

Population	**Espolla**		**Königswartha**		
Sexual System	Dioecious		Androdioecious		
Sex	Female	Male	Hermaphrodite		Male
			Monogenic	Amphigenic	
Sex Chromosome Composition	ZW	ZZ	WW	ZW	ZZ
Sex Linked SNPs (Approach 1 & 2)					
Expected allele frequency of reference allele	0.5	1	0*	0.5	1
Zygosity	Hetero-zygous	Homo-zygous	Homo-zygous*	Hetero-zygous	Homo-zygous
W Specific RAD Tags (Approach 3)					
Normalised Coverage	0.5	0	1	0.5	0
Presence/Absence (+/-)	+	-	+	+	-
Z Specific RAD Tags (Approach 3 & 4)					
Normalised Coverage	0.5	1	0	0.5	1
Presence/Absence (+/-)	+	++	-	+	++

Polymorphisms in restriction sites

Due to the nature of RAD data, indels or SNPs in the restriction site of a particular sex chromosome could also give rise to W/Z specific RAD tags. To investigate restriction site polymorphisms, sex specific RAD tags were translated, using 'the dictionary' (see 'cross-library comparisons'), to identify the homologue from the other assembly. Using the coordinates of the translated RAD tag the restriction site was viewed using an integrative genomics viewer. This then allowed us to determine whether the corresponding gametolog was present in the other assembly and identify the causation of sex specificity, allele dropout or sequence divergence.

Monogenic/ Amphigenic Identification in androdioecious population

In androdioecious populations, hermaphroditic *T. cancriformis* can be monogenic (WW) or amphigenic (ZW). To satisfy the criteria of the sex linked SNPs investigation the monogenic hermaphrodites were identified and excluded from the dataset before applying approaches 1 and 2. Afterwards, we confirmed the status of the identified sex-linked SNPs by checking that the monogenics were homozygous for a different allele than the males.

To identify monogenic individuals, the normalised depth coverage of a putative W specific allele was plotted against that of a putative Z specific allele for all individuals. W alleles were first revealed through the 'absence/presence approach', present in all hermaphrodites and absent from males, in the Königswartha population. The genomic coordinates of resulting W specific alleles were then cross referenced with the 'catalog' to identify another allele in a relatively similar location on the genome which was present in all males (putative Z specific allele). The normalised coverages of these alleles were then plotted against each other. It would be expected that the normalised coverages would differ for the individuals depending on the monogenic or amphigenic nature of the sex chromosomes, for example, an amphigenic (ZW) individual would be expected to have a normalised coverage of 0.5x for both alleles (Table 2).

The four approaches described above were implemented using custom UNIX scripts and Excel filtering (Supplementary, Figure 9.5). Approaches 1 and 2 were performed on the outputted VCF files, containing information on SNPs within paired-end reads across a population, from Freebayes only. Approaches 3 and 4 were performed on

the "***.haplotype.tsv" files, containing bi-allelic variant information within the first-end only RAD tags from the catalog, outputted by Stacks. Approaches 1 and 2 approaches were performed for first end reads and paired end reads separately to assess the additional value of utilising the second end reads and this was also repeated using both genome assemblies to map the RADseq data (Male assembly and Hermaphrodite assembly).

Cross-catalog comparisons

As each approach used generates different codes for each RAD tag, a 'dictionary' of RAD markers was created to enable the comparison of results (Supplementary, Figure 7). We compared the putative sex linked alleles and sex-specific RAD tags found in this study with those found by Mathers et al. (2016). Consensus RAD tags outputted by Stacks using the hermaphrodite and male reference genomes were separately compared against the consensus sequences from Mathers et al. 2016 (http://datadryad.org/resource/doi:10.5061/dryad.4635t) using BLASTn from the BLAST suite (Altschul et al., 1990). The results of each BLAST were then added to a database to provide a direct translation between corresponding male reference genome RAD tags, hermaphrodite reference genome RAD tags and no reference genome RAD tags from Mathers et al. (2016).

Sex Linkage Analysis

Characterisation of sex-linked contigs

Recombination suppression (either by chromosomal rearrangements, heterochiasmia or other factors) is expected to increase linkage disequilibrium in sex chromosomes (Ostberg et al. 2013). To compare the pattern of linkage disequilibrium of sex-linked SNPs versus the general genome-wide linkage we plotted a pairwise comparison of SNPs from each population in the same contig against physical distance between SNPs in a linkage disequilibrium decay plot. It is expected that as physical distance between SNPs increases, population linkage disequilibrium will decrease. SNPs exhibiting complete linkage disequilibrium, despite large distances, will suggest that these SNPs are in regions under recombination suppression. Through the use of paired end sequence data, the possible maximum physical distance between SNPs in

the same RAD tag is increased, when compared to single end data. We used this method to investigate if sex-linked regions may be under decreased recombination suppression as well as identifying regions of the genome in full linkage disequilibrium but not conforming to a sex linked pattern.

Approaches 1 and 2 identified SNPs that are putatively sex linked but patterns were investigated throughout the whole contig through the linkage disequilibrium plot. SNPs that are in full linkage disequilibrium throughout the contig will be defined as fully sex linked with expected complete recombination suppression. Contigs that harbour sex linked SNPs but do not appear as fully sex linked were further investigated to determine whether they are partially sex linked, one or more SNPs adhering to sex linked patterns (e.g. if contig is subjected to relaxed recombination suppression in the PAR), or if the SNPs exhibiting expected sex linked patterns are false positives.

We used PLINK version 1.9 (Purcell et al. 2007) to conduct the pairwise comparison of linkage disequilibrium between SNPs on each contig, utilising the 'Pairwise LD' commands. This was performed on each population, analysing SNPs present in all individuals ≥ 20x coverage, the physical distance between each pairwise comparison was calculated through a custom UNIX script (Supplementary, Figure 9.7) before being log transformed and concatenated to the R^2 values used to measure linkage disequilibrium outputted by PLINK. The linkage disequilibrium decay plot was then created in R (R Core Team, 2016).

Haplotype Network Analysis

Haplotype Phasing

Through the comparison of sequences in the putative sex-linked contigs, the pattern of evolution of sex chromosome haplotypes in both populations can be inferred. By generating minimum spanning networks analysing relatedness between sex-linked haplotypes the histories of the populations can be better understood. For the haplotype analysis we used all SNPs within putatively sex linked contigs in both populations, Espolla and Königswartha. The VCF files outputted by Freebayes contained necessary information such as base position, base identification and genotype of the SNPs, but to establish which SNPs were physically linked within a chromosome required the SNPs to be 'phased'. ShapeIt2 (Delaneau et al., 2013) is a program that

estimates which chromosome a particular SNP is present on, allowing us to infer which SNPs within a single scaffold are physically linked. VCF files of all putatively sex linked scaffolds were produced, listing SNPs with a minimum quality of base calling equivalent to 99%.

A custom script (Supplementary, Figure 9.6) was produced using the command --*extract-PIRS* and a 'BAM list' (linking individuals to raw mapped reads) allows us to increase the quality of the phase prediction by utilising 'phase informative reads' (PIRs). Raw sequencing reads spanning at least two heterozygous sites (PIRS) will be read from BAM files to increase the power of the phase estimation for SNPs with coverage depth ≥ 20x. These PIRs were incorporated into ShapeIt2 command, outputting files containing haplotype information and sample information. Using a custom Python script (Supplementary, Figure 9.7), these output files were translated into FASTA files then trimmed and and aligned manually using AliView (Larsson, 2014). Indels were replaced with a random non-occurring base to signify one mutation/nucleotide change.

Haplotype Network

Haplotype networks were produced for all scaffolds that contained at least one sex linked SNP in both populations. The R package *pegas* was used to build the trees that connect haplotypes based on pairwise differences.

The haplotypes of each phased contig were assigned to a particular sex chromosome in each population through their presence within a particular sex and the expected sex chromosomes to be found in them. The following protocol was devised:

- A haplotype found only in Königswartha monogenic hermaphrodites (WW) as well as either Espolla females (ZW) or Königswartha monogenic hermaphrodites (ZW) was assigned to a W chromosome.
- Haplotypes found in Espolla males (ZZ) or Königswartha males (ZZ) in addition to Espolla females (ZW) or Königswartha amphigenic hermaphrodites (ZW) were characterised as a Z chromosome.
- Some haplotypes are rare and only found in a small number of individuals. Therefore, if a haplotype is found in a single sex then it was characterised as a Z chromosome if it was only found in males (ZZ) or a W chromosome if it was

only found in monogenic hermaphrodites (WW). Haplotypes found in just amphigenic hermaphrodites (ZW), Espolla females (ZW) or a combination of both were assigned to Z if they most closely resembled already identified Z haplotypes, if not they were assigned to W.

- If a haplotype is found in both males (ZZ) and monogenic hermaphrodites (WW) it is characterised as not sex linked.

After haplotype networks had been analysed, all contigs that are putatively fully sex linked in both populations were concatenated into a 'superhaplotype,' assuming physical linkage between the sex-linked contigs. The contigs were phased using a 'maximum likelihood' approach and merged together. By assuming physical linkage between contigs, we assume they all belong to the non-recombining region of a single sex-chromosome, from this we can determine the structure of the network of the sex chromosomes. This method cannot include chromosome specific alleles without a specific counterpart, and so is restricted to measure the structure of the least divergent areas of the sex chromosomes.

Estimation of the size of the sex linked region

To determine the size of region in the sex chromosomes where recombination is suppressed the base pair length of fully sex-linked and W and Z specific contigs from the reference genomes were extracted from the metadata in the VCF files. The lengths of the putatively sex linked and sex-specific contigs within and among both populations were then added to determine the putative size of the sex linked region in each population and the size of the sex linked region shared between populations.

Only fully sex linked contigs identified by the linkage disequilibrium analysis and contigs containing sex specific RAD tags with no evidence of polymorphisms in the restriction site. If a sex specific RAD tag had a polymorphism in the restriction site it was assumed that it is fully sex linked in the upper bound, as we could not confirm that all of the SNPs within the RAD tag or contig were fully sex linked, giving bias to SNPs in the restriction site. The size of the W specific region, Z specific region and sex linked region was determined for each population and between populations as size of regions may differ between populations.

Identification of panel of sex-linked genes and functionality

Genome assembly annotation and gene prediction

The genome assemblies separately underwent structural annotation and gene prediction using the Maker2 pipeline (Holt and Yandell, 2011). Repeatmodeler (Smit and Hubley, 2015) and Repeatmasker (Smit, Hubley and Green, 2015) were used to identify repeat regions of the assembly, after this, CEGMA (Parra et al., 2007) was used to train SNAP, an *ab initio* gene predictor software (Korf, 2004). Maker2 underwent a first run to determine gene models whilst accounting for repetitive regions, predictions from SNAP and information from physical evidence. Physical evidence was composed of Swiss-prot curated proteins (The UniProt Consortium, 2017), short single-read transcript *Triops* sequences extracted from the 'EST' database from GenBank (Benson et al., 2005) combined with *Triops newberryi* transcriptome (Horn et al., 2017). The output gene models from this first pass of Maker were used to retain the SNAP software and train another *ab initio* software, Augustus (Stanke and Morgenstern, 2005). The profiles of theses two gene predictor softwares were utilised in a second run in Maker2.

The gene models outputted from the second run of Maker were functionally annotated through comparison against a metazoan subset of NCBI's nonredundant (nr) protein database using BLASTP (Altschul et al., 1990) and also the Interproscan pipeline (Jones et al., 2014), which uses sequences to be scanned against predictive models from a number of different databases. BLAST2GO (Götz et al., 2008) was then utilised to compile the results from the comparisons and infer a consensus functional annotation and corresponding GO terms.

Identification of candidate sex-linked genes

A database of gene descriptions comprising all scaffolds in each assembly, outputted from BLAST2GO, was filtered to only include scaffolds including sex linked RAD. Each gene description from this database was then searched against the gene database from UniProt, recording any gene aliases found in *Drosophila melanogaster*. *D. melanogaster* was used as it provided a compromise of most comprehensive gene function analysis and relatedness to *Triops cancriformis.* If no results for *D. melanogaster* were found, genes from other model organisms such as *Homo sapiens*

or *Mus musculus* were used and their corresponding ortholog derived using the g:Orth function from g:Profiler (Reimand et al., 2007).

In order to identify sex-linked genes with functions related to sex determination or potential sex-antagonistic selection, the resulting *D. melanogaster* genes were then analysed using the g:GOSt function from g:Profiler, with *D. melanogaster* (Ensembl) (Reimand et al., 2007) as the reference genome. Gene ontologies terms were then mined for GO terms associated with sex determination using a defined term list: “SEX” , “MALE”, “FEMALE”, “HERMAPHRODITE”, “SEXUAL”, “OOCYTE”, “SPERM”, “TESTES”, “OVARIES”, “SEXUAL”. As there is no fully annotated *Triops cancriformis* reference genome in Ensemble that outputs of this test are limited, therefore only biological processes and umbrella terms at a GO level of 5 or below were utilised. A literature search on PubMed and Flybase was performed to support putative functions.

Results

Demultiplexing, Quality control and Mapping

A total of 228,609,566 reads were obtained from the combined four libraries, after filtering for low quality or error prone reads and truncation through *process_radtags* a total of 217,927,985 reads were retained, composed of 105,333,023 read pairs. Following the removal of PCR duplicates via the *clone_filter* command a total of 27,865,478 read pairs were retained, a 24.38% of the the total raw (Table 3).

Table 3. Read information retained after the demultiplexing process using Stacks (Catchen et al., 2013). Reads retained were used in downstream analysis.

	Process_radtags		Clone_filter		Overall
Library	**Input Reads**	**Retained Reads**	**Input Read Pairs**	**Read Pairs Retained**	**Retained Reads (%)**
Library 1	51,328,706	49,007,250	23,740,812	6,304,763	24.57
Library 2	72,216,624	68,960,128	33,347,549	9,160,859	25.37
Library 3	47,001,616	44,631,465	21,507,737	7,333,676	31.21
Library 4	58,062,620	55,329,142	26,736,925	5,066,180	17.45
Total	228,609,566	217,927,985	105,333,023	27,865,478	24.38

The filtered and cleaned RAD sequence data was mapped to both male and hermaphrodite reference genomes separately. It would be expected that if there was significantly diverged regions of the sex chromosomes this would result in an increased mapping of the Z chromosome RAD tags to the male assembly (ZZ) and an increased mapping of the W chromosome RAD tags in the hermaphrodite assembly (WW), as diverged regions would not map. Despite this, we noticed very few differences in the mapping between the two assemblies (Appendix, Figure 2). The percentage of reads not hitting any loci on the assembly was highly variable between individuals on both assemblies, varying from an average of 59.44% in ESP_F11 to 5.64% in ESP_M5 between both assemblies, although the proportion of hits on multiple regions to unique hits was stable across individuals (Supplementary – Figure 10). The high frequency of ‘no hits’ in some individuals is consistent with the high levels of contamination detected by Mathers et al. (2015).

RAD tag mapping and SNP calling

Stacks *ref_map.pl* generated a total of 1733 and 1725 RAD tags using the male and hermaphrodite assembly mapped data respectively, with respective average coverage depths of 294.63x and 294.66x (Appendix, Figure 1). From the 1733 RAD tags utilising the male assembly, 719 were present in all individuals with no variation between individuals, representing putatively autosomal RAD tags, and 793 from the 1725 RAD tags obtained using the hermaphrodite assembly were putatively autosomal.

Table 4. Summary of SNPs called when using first end reads only in comparison to using paired end reads (first and second reads) across the two populations using the two different genome assemblies

Reference Assembly	**Male**				**Hermaphrodite**			
Population	Espolla		Königswartha		Espolla		Königswartha	
RAD tags* in catalog	1733				1725			
Read type	First End	Paired End	First End	Paired End	First End	Paired End	First End	Paired End
SNPs Called	500	3696	117	997	536	4446	157	971
Additional SNPs	3196		880		3910		814	

SNP calling was performed both on forward reads only and on paired end reads (forward and reverse reads combined) to assess the increase in data obtained by utilising the paired end reads. VCF files containing SNPs were filtered to only those present in >75% of individuals, minor allele frequency of > 0.05 and coverage depth ≥ 20x. The number of SNPs called increased when utilising the reverse end reads in conjunction with forward reads. On average, 7.72x more SNPs were called when using the forward and reverse reads combined (Table 4).

Population polymorphism and differentiation

Using the male assembly RAD tags present in all individuals of at least one population, F_{ST} between the populations was 0.726, this is close to the figures of 0.710 reported by Mathers et al. 2015 and 0.719 reported by Zierold et al., 2009.

SNP calling shows that *T. cancriformis* population from Espolla is more polymorphic than Königswartha, for example, using forward reads only, we found an average SNP calling rate of 4.28 SNPs per kb across the Espolla genome (averaging between both assemblies), in comparison, an average of only 1.13 SNPs per kb called in Königswartha, again between both assemblies. This pattern persisted using when using the second end reads (Table 4).

Differences in the sex-linked region between populations and sexual systems

Sex linked RAD tags

After merging initially segregated indel-possessing alleles, we found that Espolla have 299 putatively sex linked SNPs residing on 43 RAD tags from 24 contigs using the male assembly and 232 putative sex linked SNPs composed from 41 RAD tags from 22 contigs using the hermaphrodite assembly. In contrast, Königswartha we only found 24 putatively sex linked SNPs on 16 RAD tags using the male assembly and 42 putatively sex linked SNPs on 22 RAD tags using the hermaphrodite assembly on 11 and 14 contigs respectively (Table 5).

One of the sex linked RAD tags found using the male assembly was actually found to be a W linked RAD tag through mapping to the male assembly, with this RAD tag (MALE_424), matching to two RAD tags in the hermaphrodite assembly (HERM_586, HERM_970) in the 'dictionary'. Through further analysis of normalised depths of this RAD tag, HERM_970, it is seen that this RAD tag has twice the expected coverage in females and hermaphrodites (Supplementary, Figure 1). This is consistent with the false mapping of reads to the assembly, wherein the females and hermaphrodites have two very similar regions on the genome but the males only one. When the mapping assembly is ZZ, reads are merged to one area, but using a WW assembly, reads are split during mapping. We can predict, from this, the reads are from a W specific RAD tag, as is present when utilising the hermaphrodite assembly, and therefore RAD tag MALE_424 was removed from all downstream analysis. After the

exclusion of MALE_424, we found that Espolla have 288 putatively sex linked SNPs residing on 42 RAD tags from 23 contigs using the male assembly (Table 5).

After production of 'the dictionary' these putatively sex linked RAD tags could be compiled and we could determine the number of unique RAD tags showing a sex linked pattern across the two populations. 49 unique RAD tags exhibited expected sex linkage patterns, 17 RAD tags from 11 contigs were present and sex linked in both populations, 26 putatively sex linked RAD tags from 15 contigs were sex linked only in Espolla and 6 putatively sex linked RAD tags from 4 contigs were only present and showed sex linkage patterns in Königswartha (Supplementary, Figure 2).

Table 5. Summary of sex linked/ specific RAD tags and contigs discovered using the adapted methods from Brelsford et al. (2016). Different assemblies are displayed, *None refers to the data obtained in the *de novo* study by Mathers et al. (2015). ** refers to data was not investigated due to divergence between reference genome and organism (Aligning Z in KOE to W in Scotland). *** refers to contig information only possible for reference mapped reads

Population	*Espolla*			
Reference Genome Assembly	Male	Hermaphrodite	Unique	*None
Sex Linkage				
W Specific RAD Tags | Contigs	6 | 5	15 | 10	17 | 12	18 | ***
Z Specific RAD Tags | Contigs	8 | 6	**	8 | 6	1 | ***
Sex Linked RAD Tags | Contigs	42 | 23	41 | 22	43 | 24	22 | ***
Population	*Königswartha*			
Reference Genome Assembly	Male	Hermaphrodite	Unique	*None
Sex Linkage				
W Specific RAD Tags | Contigs	1 | 1	2 | 2	2 | 2	2 | ***
Z Specific RAD Tags | Contigs	3 | 3	3 | 2	4 | 3	11 | ***
Sex Linked RAD Tags | Contigs	16 | 11	22 | 14	23 | 14	3 | ***

Sex specific RAD tags

W specific RAD tags

Espolla

After filtering for spurious indel splitting from Stacks, the 'catalog' filtering for the Espolla RAD tags resulted in 15 RAD tags from 10 different contigs when using the hermaphrodite assembly to map the data (Table 5). Using the male assembly to map the data resulted in 6 W specific RAD tags from 5 contigs, after indels filtering (Table 5). The 'dictionary' was then used to determine the unique number of W specific RAD tags across the two assemblies. Overall we found 17 unique W specific RAD tags found within 12 contigs: with 4 RAD tags present using both assemblies, 11 RAD tags present using only the hermaphrodite assembly and 2 being present using only the male assembly (Figure 2a, 2b). Nine out of the 17 unique W specific RAD tags were a result of polymorphisms in the restriction site causing null allele dropout, with an equivalent gametolog Z specific allele in the male genome assembly (Supplementary, Figure 2).

Königswartha

After indel removal just two RAD tags from two contigs were found to be W specific in Königswartha when mapping to the hermaphrodite assembly (Table 5). Filtering the 'catalog' from the male assembly mapped data for W specific RAD tags resulted in one RAD tag from one contig when accounting for indels. Combining the W specific RAD tags from the two assemblies resulted in two unique W specific RAD tags from two contigs. However, one of these RAD tags was had much lower expected normalised coverage for a W-specific RAD tag (Table 2), so was excluded from downstream analysis (Figure 2c, 2d).

Identification of monogenic hermaphrodite individuals

Prior to our analysis of sex linked alleles/SNPs in Königswartha, the monogenic individuals needed to be extracted from our data set. To determine which of our individuals were monogenic (WW), a putatively W specific allele (MALE_1020) was identified with a putatively Z specific allele (MALE_1116) located 3 bp away (this was determined to be a sex linked indel split by Stacks). By plotting the coverage of the

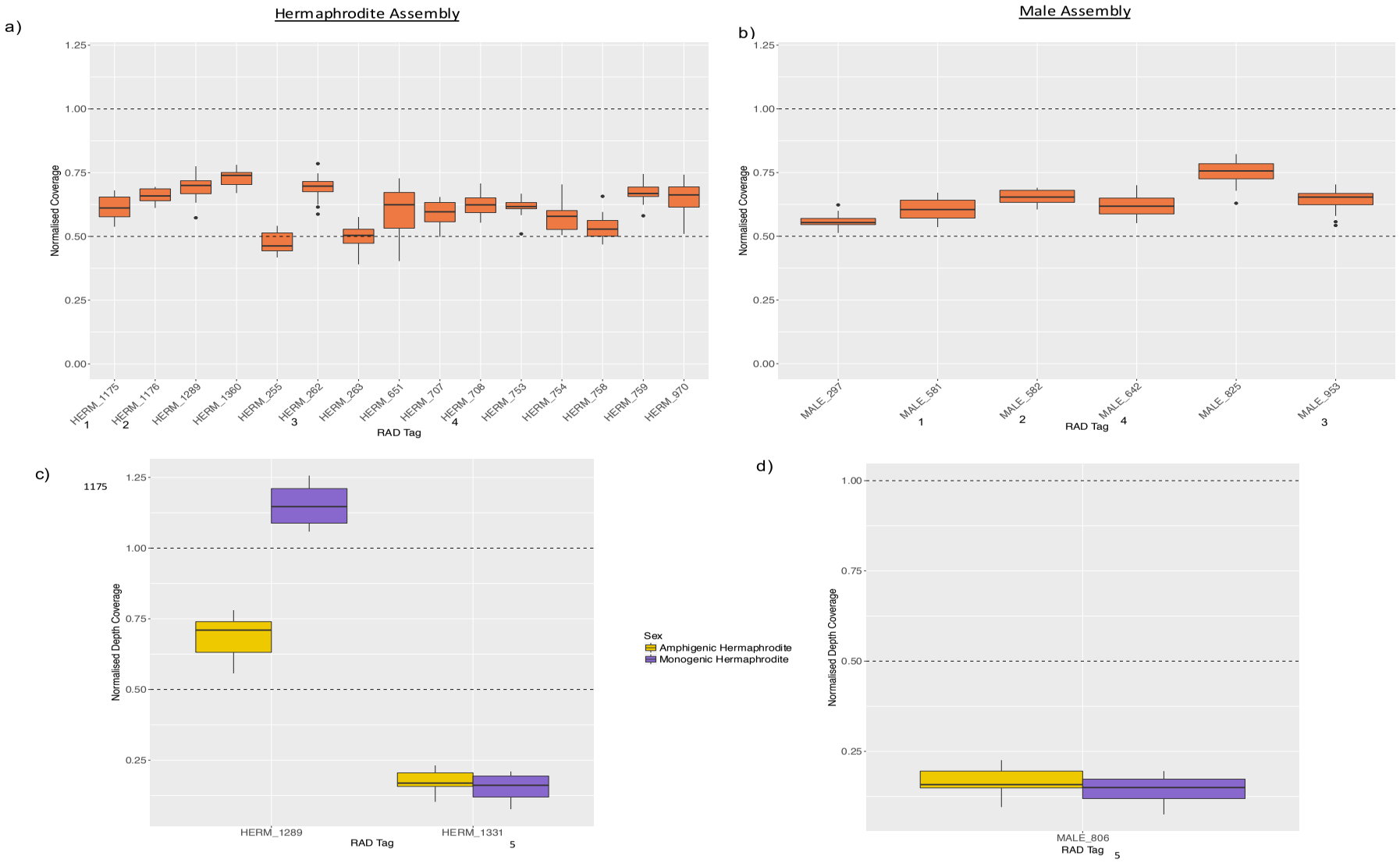


Figure 2. The normalised depths of RAD tags satisfying the pattern of expectations of a putative W specific RAD tag. Figures a and b show the normalised depths for Espolla population whereas c and d are from the Königswartha population. Corresponding numbers indicate cross-catalog RAD tags comparedthrough 'the dictionary'. HERM_1331/MALE_806 has a much lower coverage than expected for a W specific RAD tag. HERM_1289 is the only RAD tag that is W specific among both populations, although HERM_970 is W specific in some individuals in Königswartha and HERM_262/MALE_953, HERM_263, MALE_297 and HERM_651 all have sex linked alleles in Königswartha.

two RAD tags against each other we concluded that 4 individuals met the criteria expected from a monogenic (WW) system: KOE_H10, KOE_H21, KOE_H25, KOE_H29 (Figure 3) in agreement with Mathers et al (2015).

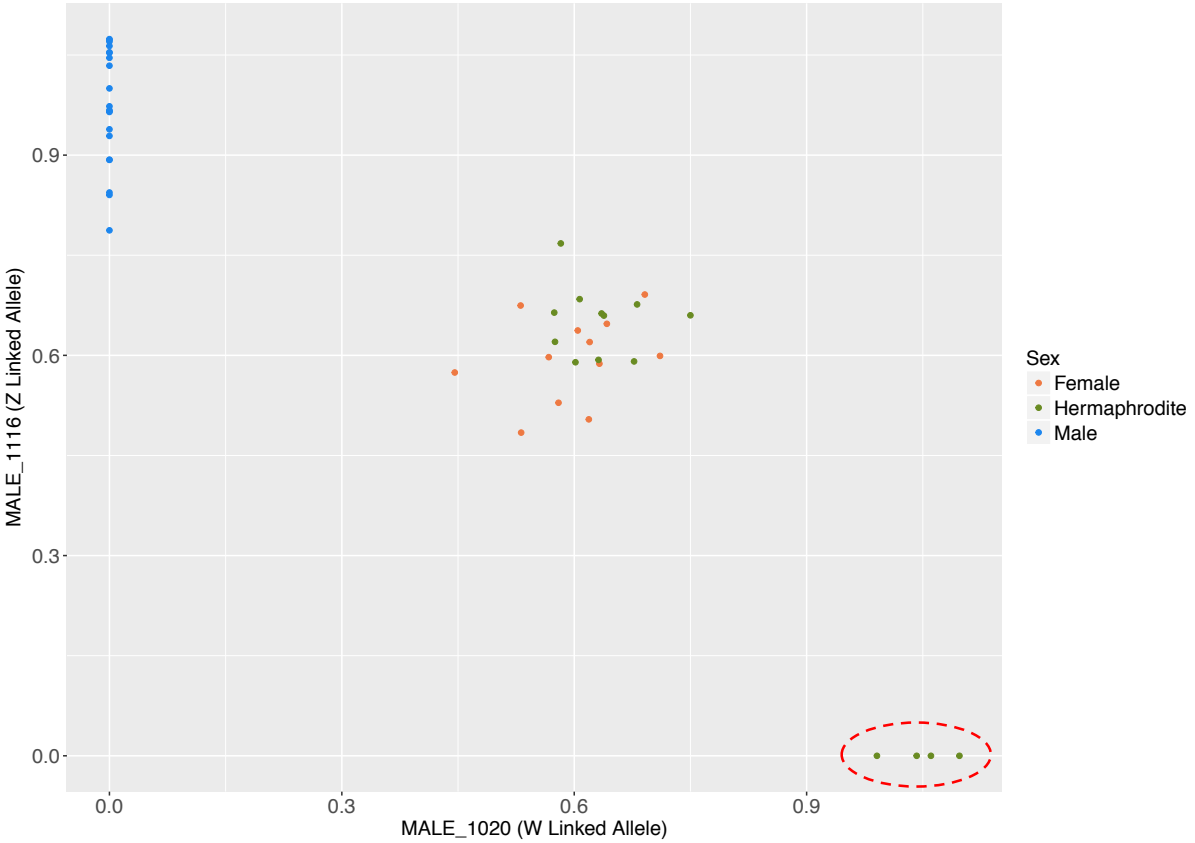


Figure 3. The identification of monogenic individuals through comparison between coverage depths in W/Z specific alleles. Four individuals adhere to expectations of WW individuals (Table 2), displayed in the red dotted circle.

Z specific RAD tags

Espolla

The plot of average normalised RAD tags coverage depths of male against females from the Espolla 'catalog' revealed a cloud of RAD tags with the expected coverage in Z linked markers (Figure 4). After discarding RAD tags split by indels, this cloud comprised eight RAD tags across six contigs when mapping to the male assembly (Table 5). Two of these Z specific RAD tags were a result of null allele dropout from polymorphisms in the restriction site (Supplementary, Figure 2).

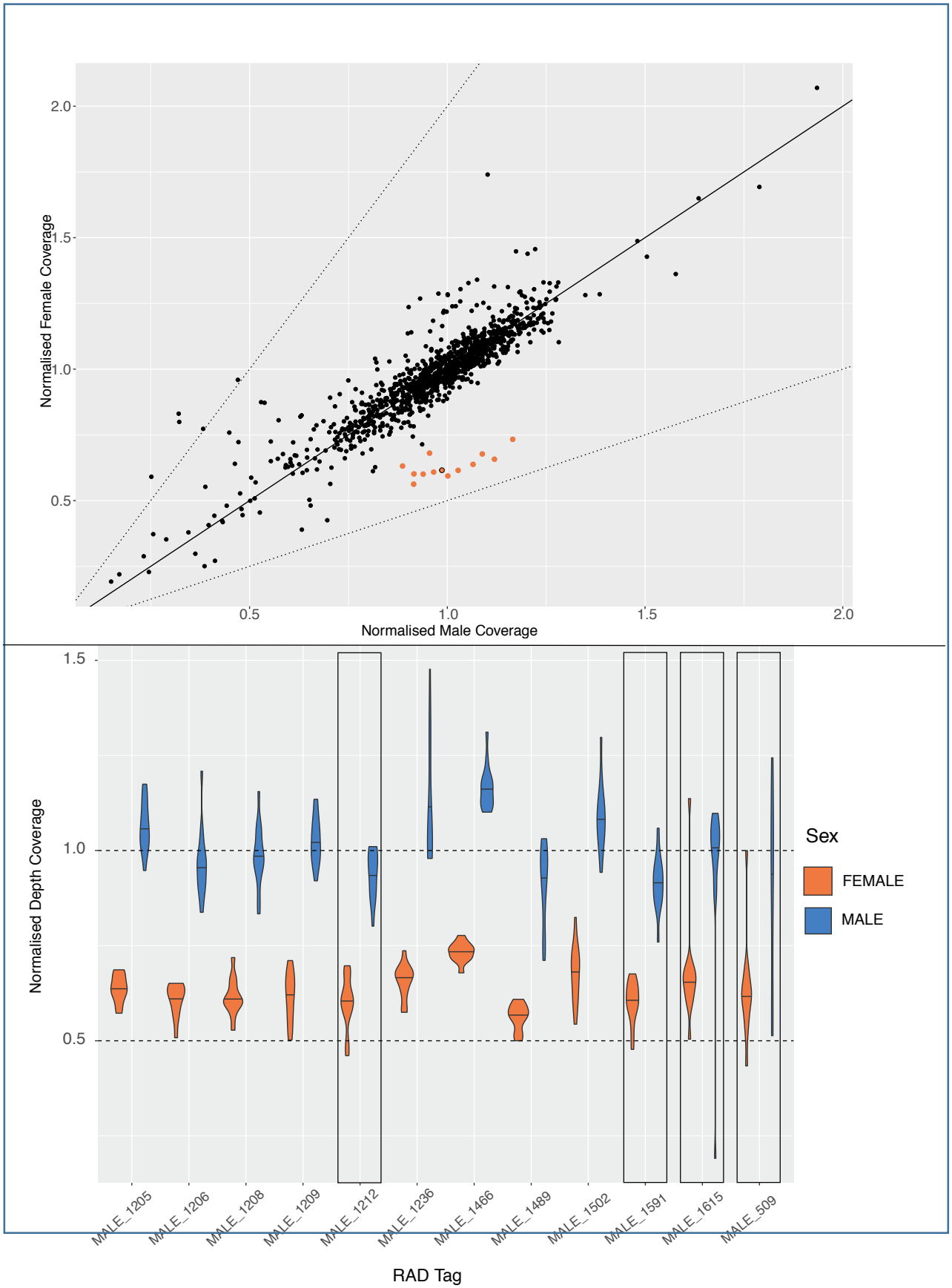


Figure 4. The average normalised depth comparisons used to determine Z specific RAD tags in Espolla. The top panel is the average normalised coverage depth plot used to identify possible Z specific RAD tags. The cloud of orange dots represents the identified putative Z specific RAD tags, the orange and black dot resembles an indel later identified by Freebayes. The bottom panel is a violin plot used to confirm patterns of coverage, RAD tags in boxes failed to meet the > 0.35 divergence between normalised male and female depths.

Königswartha

Filtering of the Königswartha 'catalog' for RAD tags present in all male and amphigenic hermaphrodite individuals but absent in monogenic hermaphrodite individuals revealed three Z specific RAD tags from two contigs after indel identification when utilising the hermaphrodite assembly. After indel removal three Z specific RAD tags from three contigs were identified using the male assembly. Through cross comparison using the 'dictionary' it was found that these represented four unique RAD tags across three contigs, but 3 of these were found to have resulted from null allele dropout due to restriction site mutations (Figure 5; Supplementary, Figure 2).

Comparison to Mathers et al., 2016

Overall we found 16 W specific RAD tags present in 11 contigs in Espolla and one RAD tag from one contig that was W specific in Königswartha. Only one W specific RAD tag, HERM_1289, was found in to be W specific in both populations although HERM_970 displayed expected W specific depths in some individuals of the Königswartha population (Supplementary, Figure 2). We found seven RAD tags present in five contigs to be Z specific in Espolla and three RAD tags from three contigs to be Z specific in Königswartha with no shared Z specific RAD tags, although one contig was shared between populations. In total, we found 50 unique sex linked RAD tags containing at least one sex linked SNP, of these 26 were sex linked in Espolla only, 6 sex linked in Königswartha only and 17 present in both populations. Of the 26 RAD tags containing only sex linked in Espolla, 15 RAD tags were present in Königswartha but did not show any sex linked patterns, whereas from the six RAD tags only sex linked in Königswartha three RAD tags were not present in Espolla and three were W specific RAD tags.

Mathers et al. (2015) found 52 RAD tags that are sex specific or linked, but through our methods we were able to increase this to 74. Mathers et al., 2016 found 18 W specific RAD tags in Espolla and two W specific RAD tags in Königswartha, in this study we found 15 W specific RAD tags in Espolla and two W specific RAD tags in Königswartha. Mathers et al. 2016 identified one Z specific RAD tag in Espolla and 11 Z specific RAD tags in Königswartha, in comparison we found seven Z specific RAD tags in Espolla but only four Z specific RAD tags in Königswartha. The largest

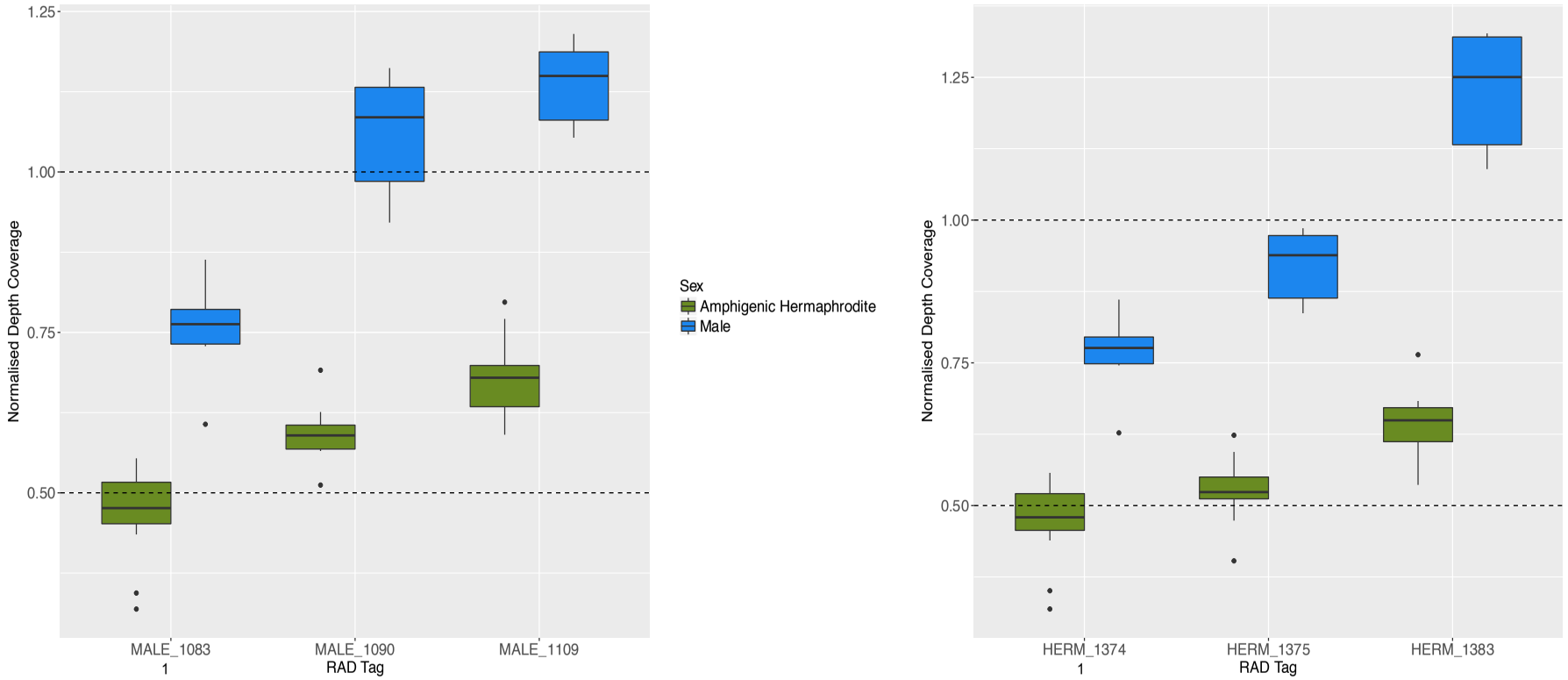


Figure 5. The RAD tags that conform to the depth coverages expected of Z specific RAD tags. Corresponding numbers indicate matching RAD tags determined through the ‘dictionary’.

difference noted between Mathers et al. (2015) and our results was in regards to the sex linked RAD tags, Mathers et al. (2015) uncovered 22 and 3 putatively sex linked RAD tags in Espolla and Königswartha respectively, whereas we found 45 putatively sex linked RAD tags in Espolla and 23 putatively sex linked RAD tags in Königswartha.

Linkage disequilibrium decay assessment on sex linked region

Espolla

The plot of linkage disequilibrium (LD) decay suggested that a total number of 191 out of 219 contigs within the Espolla genome displayed full LD between at least two SNPs, with genomic distances up to 68 kb (Figure 6; Supplementary, Figure 8). All SNPs from 22 out of 24 contigs possessing at least one sex linked SNP displayed an R^2 value of 1, that is, were fully linked throughout the contig.

The only exceptions were two contigs containing at least one sex linked SNP but did not display full LD across the contig, *'N_1073'* and '*N_1083'* (Supplementary, Figure 3). The contig *'N_1073'* had 14 SNPs, 13 of which were sex linked and in full LD. However, one SNP seemed to cause deviation from full linkage, females carry either of the two genotypes CT/CC, whereas males possessed any of three different genotypes, CT/CC/TT (Supplementary, Figure 3). This SNP has high confidence of accuracy and a depth of coverage expected in autosomal loci, indicating it is not a result of a duplication event. Males have two alleles present in either chromosome, whereas females have the two alleles present in only one chromosome, this pattern is consistent with what we would expect from a reverse mutation in the Z chromosome. In the contig '*N_1083'* only one out of nine SNPs are conforming to an expected sex linked pattern, this suggests that the SNP possessing expected sex linkage is either a false positive, or is partially sex linked (Table 6).

The 22 fully sex linked regions contrast with the rest of genomic regions, where a decay in linkage with distance is appreciable. Although, across the whole genome, only five contigs possessed full LD over genetic distances > 1 kb. Two of these contigs were harbouring SNPs identified as sex linked through our sex linkage identification process ('*N_187*', '*N_359*') although three contigs, *'N_571'*, '*N_783'* and '*N_3019'*, did not possess any sex linked SNPs identified by our detection methods.

Overall we have determined that 22 out of 24 putative sex linked contigs initially identified by approaches 1 and 2 are fully sex linked contigs, with '*N_1073*' contig being partially sex linked and having two possible genotypes for the Z chromosome at one SNP and '*N_1083*' being partially sex linked or not sex-linked.

Table 6. Summary of SNPs that are causing deviation from full linkage in contig with sex linked SNP. Group one consists of KOE_H10, KOE_H19, KOE_H10, KOE_H2, KOE_H20, KOE_H24 and KOE_H4; Group two consists of KOE_H18, KOE_H21, KOE_H23, KOE_H25, KOE_H26, KOE_H27, KOE_H28, KOE_H29 and KOE_H30.

Chromosome	Position	Individuals differing from surrounding SNPs	Conclusion
Espolla			
N_1073	14908	ESP_F17, ESP_F2, ESP_F20, ESP_F8, ESP_M1, ESP_M10, ESP_M11, ESP_M12, ESP_M2, ESP_M3, ESP_M7, ESP_M8, ESP_M9	Polymorphism in Z chromosome
N_1083	19013	Only SNP SL	Partially linked
Königswartha			
N_305	40379	Group 1	Hermaphrodites share male genotype
N_305	40672	No sex linkage	Partially linked
N_359	9478	Group 2	Hermaphrodites share male genotype
N_596	26141	Group 2	Hermaphrodites share male genotype
N_1198	7042	Group 2	Hermaphrodites share male genotype
N_1837	14860	Group 2	Hermaphrodites share male genotype
N_2759	3279	Monogenic hermaphrodites	Partially linked
N_2759	3426	Group 1	Hermaphrodites share male genotype
N_2759	3644	Group 2	Hermaphrodites share male genotype
N_4227	3742	Group 2	Hermaphrodites share male genotype
N_4227	3753	Monogenic hermaphrodites	Partially linked

Königswartha

From the LD decay plot 79 out of 105 contigs possessing at least two SNPs from the Königswartha population had full linkage between at least one pair of SNPs on the contig, up to 99 kb (Figure 6). All SNPs from four (*N_746, N_855, N_1093, N_1161)* out the 11 contigs that harbour ≥1 sex linked SNP have full linkage present across the contig, suggesting that these are fully sex linked contigs.

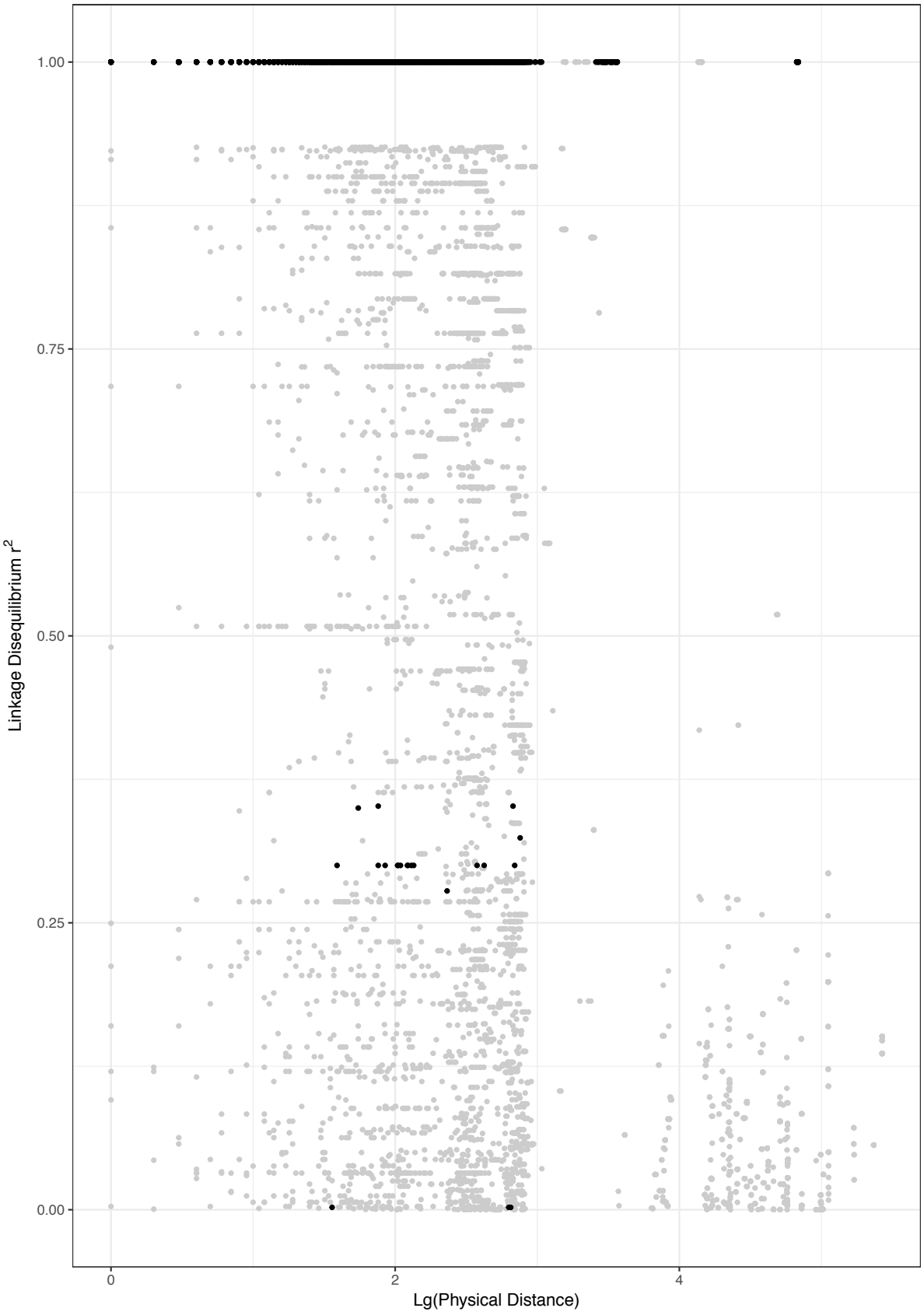


Figure 6. Linkage decay over physical distance in Espolla. Black dots refer to pairwise comparision including ≥ 1 sex linked SNP. The SNPs not fully linked are caused by pairwise comparisons between non-sex linked and sex linked SNPs in two contigs (*N_1073* and *N_1083*).

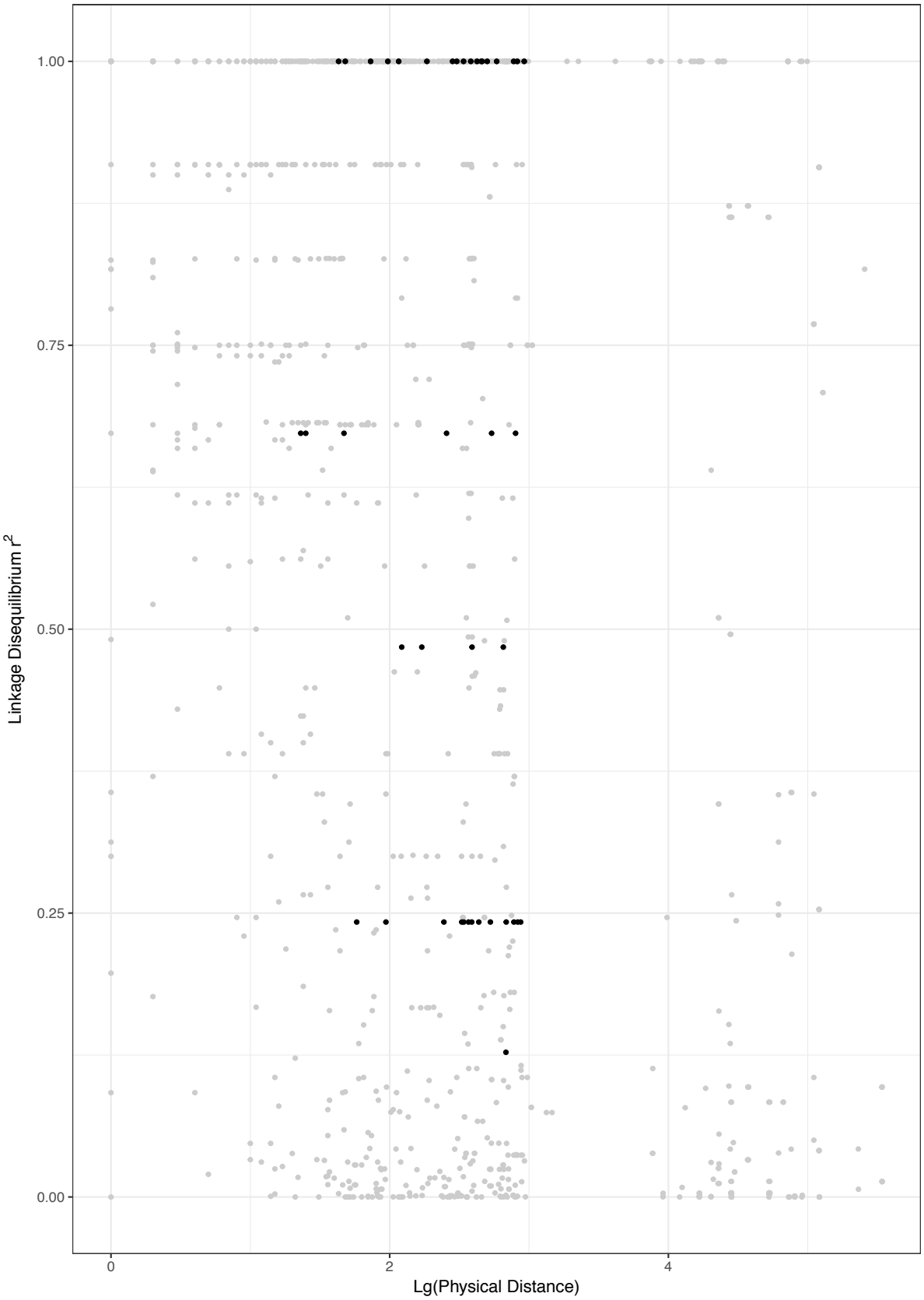


Figure 7. decay over physical distance in Königswartha. Black dots refer to pairwise comparision including ≥ 1 sex linked SNP. The SNPs not fully linked are caused by pairwise comparisons between non-sex linked and sex linked SNPs in table 6.

From the seven contigs that harboured ≥1 sex linked SNP it was determined that the deviation from full linkage across the contig was being caused by either SNPs in which there was differences between two groups of hermaphroditic individuals, with each particular group sharing a genotype with males (shared Z genotype) or only partial linkage of a contig with only a small number of SNPs being sex linked (partially linked).

Six different contigs all possessed an R^2 value of 0.24 for some pairwise comparisons, through further analysis it was observed at these loci differences were seen between two groups of hermaphrodites. At these loci five out of 11 amphigenic hermaphrodites and one of four monogenic hermaphrodites were adhering to expected heterozygosity sex linked patterns, but the other hermaphroditic individuals had identical alleles to the males at these loci (Supplementary, Figure 3). At one loci in one of these six contigs (*N_2759*) the reverse was also observed causing an R^2 value of 0.48 (Table 6). This pattern is consistent with what we would expect to see from two differing haplotypes present in a W chromosome as the differences only occur in individuals possessing a W chromosome, although this cannot be confirmed without chromosome phasing (see haplotype network analysis).

The contig *N_305* had only one sex-linked SNP and when comparing this to another SNP has an R^2 value of 0.48, with a group of hermaphrodites sharing a genotype with males (Table 6). But the other SNP on this contig had no indication of sex linkage, being stringent we must determine this contig to be partially sex linked. Contigs *N_2759* and *N_4227* appear to be partially linked contigs (Table 6), with all hermaphrodites, including the monogenic individuals characterised by heterozygous alleles at these loci producing an R^2 value of 0.67.

The fully sex linked regions contrast with a decay in linkage with distance appreciable in the rest of genomic regions, although throughout the whole genome 13 contigs possessed full LD over genetic distances > 1 kb. None of these contigs were harbouring previously identified sex linked SNPs.

Overall in Königswartha, we identified four of the 11 sex linked contigs harbouring at least one sex linked SNP to be fully sex linked contigs (*N_746, N_855, N_1093, N_1161*), with another four contigs to be fully sex linked but showing evidence of possible recombination in the W chromosome (*N_359, N_596, N_1198, N_1837*). *N_305*, *N_2759*, *N_4227* were determined to be putatively partially sex linked.

Sex linked contigs and estimation of the size of non-recombining region in both populations

For our putatively fully sex-linked region estimates we characterised any contigs which harboured only sex specific RAD tags from either assembly which were not a consequence of restriction site polymorphism dropout found from either populations, in addition to any fully sex linked contigs (clarified as such after the LD decay analysis from the male assembly). Upper bound estimates for sex linked regions include the contigs that were identified as sex specific but as a consequence of null allele dropout.

Table 7. Estimations of size of region on sex chromosome void of recombination. The size of the region is determined by the sum of fully sex linked/specific contigs. The figures in the brackets are assuming that the sex specific contigs resulting from null allele dropout are fully sex linked contigs.

Putative non-recombining region per chromosome	**Espolla (kb)**	**Königswartha (kb)**	**Shared (kb)**
W Chromosome	1,404.73 – (1,739.65)	327.44 – (370.72)	327.44 – (370.72)
Z Chromosome	1,204.21 – (1,539.13)	324.25 – (367.54)	268.14 – (311.43)
W Specific	234.93	59.29	59.29
Z Specific	34.41	56.11	0

20 putatively fully sex linked contigs were characterised as the sex linked region in Espolla, whereas six putatively fully sex linked unique contigs were characterised as the sex linked region in Königswartha, with all six of these contigs being shared between the populations. After dismissing sex specific RAD tags with polymorphisms in the restriction site, five contigs were W specific in Espolla and one contig W specific in Königswartha, this one contig was shared between populations. No contigs were Z specific in both populations, two contigs were Z specific in Espolla and one contig was Z specific in Königswartha (Supplementary, Figure 4).

Assuming that all sex-linked and sex-specific contigs identified are part of a single fully-linked region in the sex chromosomes of *T. cancriformis* genome, and that they appear as separate contigs due to assembly gaps, we estimate a much larger non-recombining region (sex specific and sex linked contigs) in the W chromosome in

Espolla in comparison to Königswartha. The entirety of the non-recombining region of the W chromosome of Königswartha as is found in the in the non-recombining region of the W chromosome of Espolla. The W specific regions are also much larger in Espolla, 234.93kb and 59.29kb in Espolla and Königswartha respectively. We estimate a larger non-recombining region of the Z chromosome in Espolla also, but in contrast to the W chromosome none of the contigs found to be Z specific in Espolla are found to be sex specific in Königswartha. One contig that is Z specific in Espolla is found to be sex linked by null allele dropout but no contigs found to Z specific in Königswartha are even found to be present in Espolla (Table 7).

Haplotype network estimates among and within populations

The diversity of the sex-linked regions and degree of differentiation between sex chromosomes across both populations was investigated through phasing and haplotyping of chromosomes within populations. For this, we used the nine fully sex linked contigs present in both Espolla and Königswartha and the 10 contigs that are fully sex linked in Espolla and present, but not sex linked, in Königswartha. There are no sex linked contigs in Königswartha that have no sex linkage in Espolla, although contig *N_746* is sex linked in Königswartha but is W specific in Espolla.

Putative sex linked contigs in both populations

Overall, SNPs phasing revealed a very low haplotype diversity of the sex chromosomes within the populations and simple haplotype networks for these contigs. The haplotype networks produced by each sex-linked contig falls into several patterns. Four of the contigs (Figure 8,a) show a single haplotype for each sex chromosome, i.e. one W chromosome and one Z chromosome. At these contigs we find that the Z chromosome haplotypes from each population are very divergent but both populations share an identical W chromosome haplotype. Three contigs, (Figure 8,b), have different W haplotypes in Königswartha, one of them, W1, is shared by Espolla and

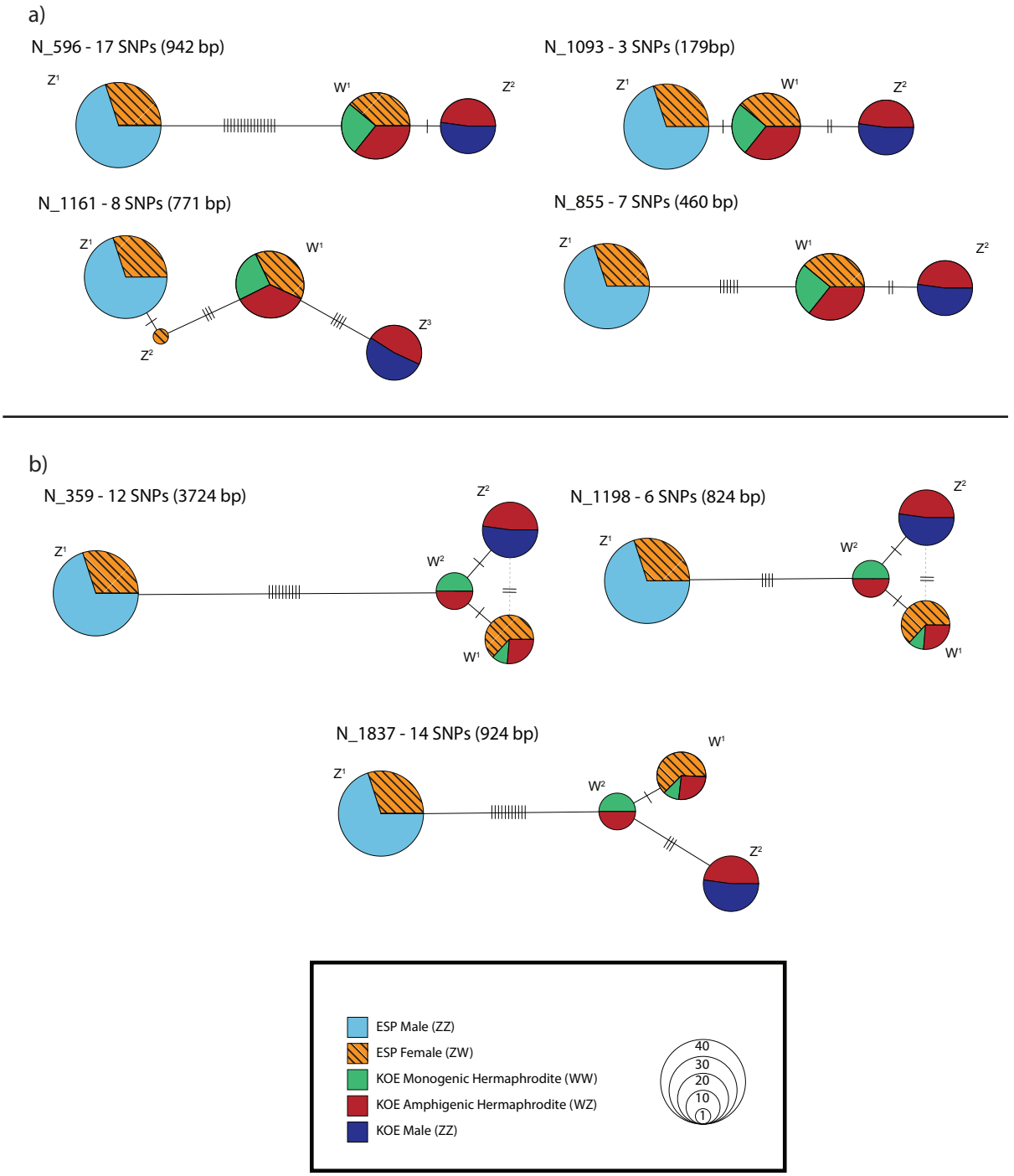


Figure 8. Networks of haplotypes present in contigs fully sex linked in both, dioecious and androdioecious, populations. Split in two different patterns: a – networks consist of a single shared W haplotype in Espolla and Königswartha with two distinct Z haplotypes, one in Espolla and one in Königswartha; b – two different Z haplotypes are present, one in Espolla and one in Königswartha but 2 W different haplotypes are present in Königswartha.

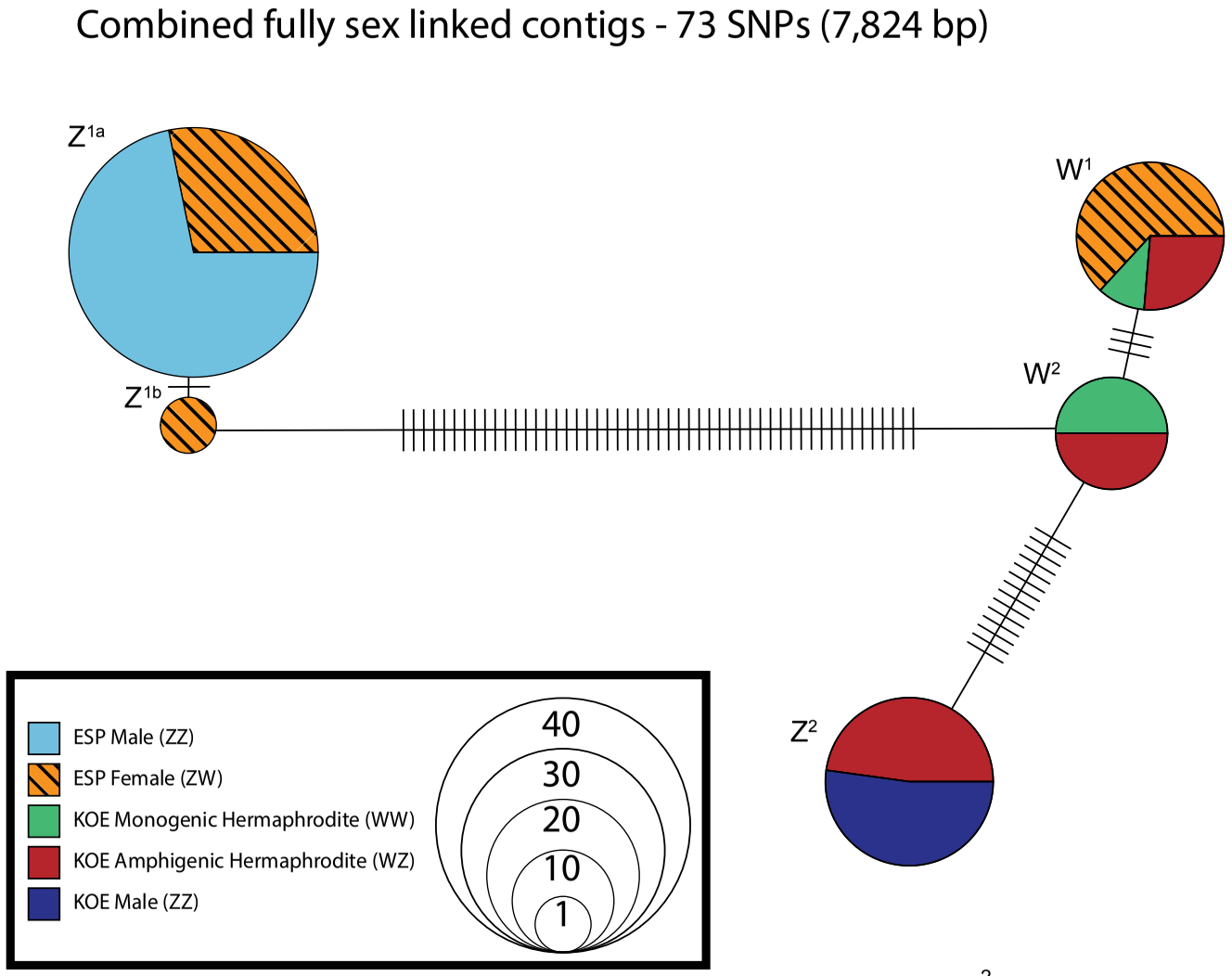


Figure 9. Overall haplotype network, assuming all fully sex linked contigs are physically linked. W^2 is located between the Z present in Königswartha and the W haplotype shared between the populations. Z chromosome present in Espolla is largely distinct from the other W haplotypes and Z present in Königswartha.

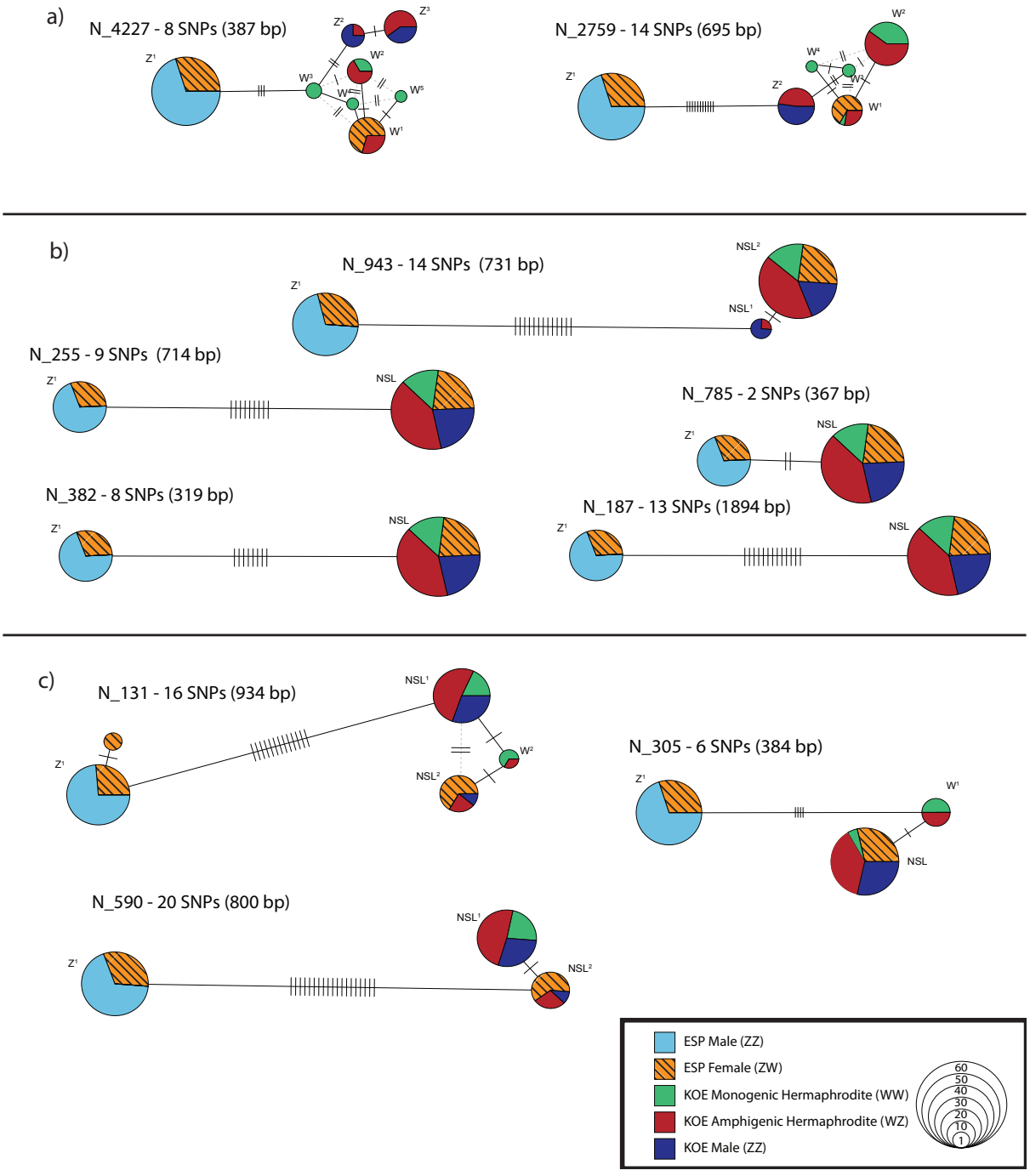


Figure 10. Haplotype network analysis of contigs that are fully sex linked in Espolla but not fully sex linked in Königswartha. Consisting of three distinctive patterns: a – deviation from sex linked patterns by polymorphisms in the monogenic individuals; b – the haplotype of the non-sex linked contig in Königswartha is identical to the W haplotype from Espolla, including in males; c – there are multiple haplotypes of the non-sex linked contig in Königswartha, one haplotype is identical to the W haplotype in Espolla and another haplotype consistently shared by a similar set of individuals.

Königswartha and is consistently found among one monogenic hermaphrodite (KOE_H10) and five amphigenic hermaphrodites (KOE_H19, KOE_H2, KOE_H20, KOE_H24, KOE_H4) whereas the other haplotype, W^2, is found consistently in three monogenic hermaphrodites (KOE_H21, KOE_H25, KOE_H29) and six amphigenic hermaphrodites (KOE_H18, KOE_H23, KOE_H26, KOE_H27, KOE_H28, KOE_H30) and is located in the network in between W^1 and the Königswartha Z, suggestive of historical recombination. Linkage disequilibrium analysis also suggested that contig N_596 would possess two different haplotypes, and therefore behave like the previous three contigs, however the SNP distinguishing both haplotypes was not genotyped in one male of Espolla and this contig was therefore omitted from the haplotype network analysis.

The different fully sex-linked contigs analysed in both populations are likely to represent non-assembled fragments of the sex chromosome. These contain 73 SNPs spanning 7,824 bp. Therefore, to assess the overall structure of the fully sex linked region in both populations, seven contigs determined to be fully sex linked in both populations (*N_359, N_596, N_855, N_1093, N_1161, N_1198, N_1837*) were concatenated and phased into a 'superhaplotype' (Figure 9). This 'superhaplotype' network shows that there are at least two very closely related Z haplotypes present in Espolla and one Z haplotype present in Königswartha. The Z haplotypes of both populations are very divergent. There is only one W haplotype present in Espolla but two different W haplotypes in Königswartha split into two groups (W^1 identical to the Espolla haplotype, and W^2 is different) (Figure 9). W^2 is located between Z^2 and W^1, although W^2 is more similar to W^1. Overall this suggests that the Z and W chromosomes are divergent in Espolla, but much more similar in Königswartha. Two W haplotypes were found within Königswartha, one of which is shared with the W from Espolla and another which is more similar to the Z from Königswartha.

Putatively fully sex linked non-sex linked contig comparison across populations

Mathers et al. (2016) determined SNPs that were sex linked in Espolla and present, but not sex linked, in Königswartha. All Königswartha individuals had a single allele which corresponded to the W allele in Espolla. Our haplotype network analysis of contigs of which are sex linked in Espolla and partially sex linked, or not sex linked in Königswartha shows that three general patterns are occurring in this region in

Königswartha (Figure 10): an expected sex linked pattern of segregation of W and Z haplotypes in Königswartha but with polymorphisms in some of the haplotypes produced by Königswartha (Figure 10, a); no polymorphism, with a single haplotype for all sexes in Königswartha, identical to that of the W haplotype from Espolla Figure 10, b); no sex linkage in Königswartha but split into two or three haplotypes with form loops suggestive of recombination (Figure 10, c).

Two contigs *N_2759* and *N_4227,* are sex linked but certain SNPs show deviation to the expected patterns as the expected homogenic individuals (ZZ males and WW monogenic hermaphrodites) are heterozygous at these SNPs (Figure 10, a). It is expected that this has been caused by the double mapping of RAD tags to a single locus, alike what was observed in *N_2269* (HERM_970). Although we can not obtain normalised coverage of the SNPs being affected by this as we can not calculate normalised coverage for second end SNPs as the depths are inconsistent. This heterozygosity has caused separate haplotypes to be produced for homogenic individuals, males and monogenic hermaphrodites in *N_2759* and *N_4227* respectively.

In a region of the putative sex chromosome that is sex linked in Espolla but not sex linked in Königswartha, five contigs retain a single haplotype in Königswartha, which is identical to the W haplotype in Espolla (Figure 10, b). Although four other contigs, are not sex linked in Königswartha despite being sex linked in Espolla have at least 2 haplotypes in Königswartha. The split between the two haplotypes is caused by polymorphisms in the individuals discovered by the linkage disequilibrium decay but with slight deviations.

Four contigs (*N_131, N_305*, *N_590, N_1073*) seem to have numerous haplotypes in Königswartha but are not sex linked (Figure 10, c). In these we see a generalised pattern of splitting of the Königswartha haplotypes for the same individuals, with few exceptions, over all four contigs. In '*N_1073'* two different haplotypes are present in the hermaphrodites of Königswartha, one of which identical to the W haplotype in the Espolla and a third haplotype in the males. Although this contig appears sex linked it contains no sex linked SNPs and does not appear to contain unusual levels of normalised coverage. In three contigs (*N_131, N_305*, *N_590*) we see a similar pattern with at least two haplotypes in Königswartha one being identical to the W in Espolla

and another different haplotype. Although the haplotypes in these contigs are caused by only one or two polymorphisms.

Candidate panel of sex linked genes

A total number of 3135 and 2436 contigs from the male and hermaphrodite assembly respectively were found to contain at least one gene annotation (Supplementary whole genome annotations), 11,579 unique genes were determined to be present across these contigs. The 69 contigs, across the two assemblies, that have at least one sex specific or sex linked RAD tag were found to contain 185 unique genes (Supplementary – Gene Aliases). Our candidate gene investigation resulted in a total of 23 genes in sex-linked contigs, from both assemblies, being identified as involved in sexual determination or sexual biological processes (Table 8).

From the male assembly, 14 genes from 13 different scaffolds were identified to be involved in sexual determination or sexual biological processes. The hermaphrodite assembly produced 18 genes from 9 different scaffolds, with 9 genes found in both assemblies (Supplementary, Figure 5). Two genes, '*Coatomer subunit alpha*' and '*Elongation of very long fatty chains acids 6*', were found in contigs possessing W-specific RAD tags in both populations whereas two genes, '*BTB POZ domain-containing protein KCTD12-like*' and '*Abl interactor*', were found in contigs that possessed W-specific RAD tags only present in Espolla. Only one gene, '*U2 small nuclear ribonucleoprotein*' was found in contigs possessing Z-specific RAD tags (Table 8).

A *Sex lethal* ortholog found in both assemblies, particularly stands out from the recovered gene panel. This gene is involved in sex determination pathways in *D. melanogaster* (Bell et al., 1988). 'Transcription factor distal-less', 'Eukaryotic translation initiation factor 4E' and 'Regulator of G-protein signalling 12' are only present in the hermaphrodite assembly but are found in regions of the genome that are sex linked in both Espolla and Königswartha, suggesting these genes are

Table 8. Gene description found in genome annotation with the alias in *D.melanogaster* with the corresponding GO term associated with the gene. Green fill denotes hermaphrodite assembly only/ yellow fill is male assembly only/ white fill is found in both assemblies.

***T.cancrifomis* Description**	**D. melanogaster ortholog**	**Sex Linkage of RADtags on Contig**	**GO Term**
Sex-lethal homolog isoform X1	SXL	Sex linked alleles ESP/KOE	Sex determination/ sex differentiation
Transcription factor distal-less	DLL	Sex linked alleles ESP/KOE	Sex differentiation
Longitudinals lacking isoforms J P Q S Z- like isoform X1	LOLA	Sex linked alleles ESP only	Sex differentiation/ development of primary sexual characteristics/ Inter-male aggression behaviour
Coatomer subunit alpha	ALPHACOP	W specific /Sex linked alleles ESP only	Sex reproduction/ male gamete generation/ male meiosis function
Metabotropic glutamate receptor 3-like	MGLUR	Sex linked alleles ESP only	Sexual reproduction
Regulator of G-protein signalling 12	LOCO	Sex linked alleles ESP/KOE	Sexual reproduction
Elongation of very long chain fatty acids 6	BALDSPOT	W specific /Sex linked alleles ESP only	Sexual reproduction
Aurora A	AURA	Sex linked alleles ESP/KOE	Sexual reproduction
Mothers against dpp	MAD	Sex linked alleles ESP/KOE	Sexual reproduction
BTB POZ domain-containing protein KCTD12-like	CG10830	W specific through allele dropout ESP	Sexual reproduction
Piwi 2	PIWI	Sex linked alleles ESP only	Sexual reproduction/ female gamete generation
Syntaxin 1A	SYX1A	Sex linked alleles ESP only	Sexual reproduction/ female gamete generation
Abl interactor	ABI	W Specific ESP/ Sex linked alleles KOE	Sexual reproduction/ female gamete generation
Juvenile hormone-inducible protein 1	JHI-1	Sex linked alleles ESP/KOE	Sexual reproduction/ female gamete generation

Major facilitator superfamily domain-containing 10	RTET	Sex linked alleles ESP only	Sexual reproduction/ female gamete generation
Transcription elongation factor SPT6	SPT6	Sex linked alleles ESP only	Sexual reproduction/ female gamete generation
Leupaxin isoform X2	PAX	Sex linked alleles ESP only	Sexual reproduction/ female gamete generation
Dorsal-ventral patterning Sog isoform X1	SOG	Sex linked alleles ESP only	Sexual reproduction/ female gamete generation
Insulin receptor substrate	CHICO	Sex linked alleles ESP/KOE	Sexual reproduction/ female gamete generation
Clathrin heavy chain	CHC	Sex linked alleles ESP/KOE	Sexual reproduction/ female gamete generation/ male gamete generation
Eukaryotic translation initiation factor 4E	EIF-4E	Sex linked alleles ESP/KOE	Sexual reproduction/ female gamete generation/ male gamete generation
U2 small nuclear ribonucleoprotein	U2A	Z Specific ESP only	Sexual reproduction/ male gamete generation
Ring canal Kelch homolog	KEL	Sex linked alleles ESP only	Female gamete generation

candidates that could potentially be involved in the sex determination of hermaphrodites or potential targets for accumulated sexually antagonistic selection. '*Longitudinals lacking isoforms J P Q S Z- like isoform X1*', '*Metabotropic glutamate receptor 3-like'*, '*Piwi 2*', '*Major facilitator superfamily domain-containing 10*', '*Transcription elongation factor SPT6*', '*Leupaxin isoform X2*', '*Dorsal-ventral patterning Sog isoform X1*', '*Syntaxin 1A'* and '*Ring canal Kelch homolog*' are all genes that are present in contigs which are fully sex linked in Espolla but not in Königswartha. Potentially, these genes may be under recombination suppression in Espolla but the recombination in this region is more relaxed in Königswartha.

During the analysis in g:Profiler, a *T. cancriformis* gene sex linked when using the male assembly only, '*NF-kappa-B inhibitor zeta*', was found to be the ortholog of the RELISH gene from *D.melanogaster*. RELISH gene has been determined to be physically associated with the male S*ex lethal* gene in *Drosophila* (Fukuyama et al., 2013), suggesting its *T. cancriformis* ortholog, 'NF-kappa-B inhibitor zeta', may also be involved in a sex determination pathway.

Discussion

We identify novel aspects of genomic architecture of sex chromosomes in *Triops cancriformis*, identifying at least 18 putatively sex linked genomic contigs, with 7 present across two populations with different sexual systems. The use of paired end RAD sequence data in combination with sequenced genomes of individuals of *T. cancriformis* allowed us to exploit second end reads, which until now have rendered impossible to confidently utilise. Through the use of paired end RAD sequencing data mapped onto draft genome male and hermaphrodite assemblies and adapted criteria for the identification of sex linked markers we were able to substantially increase the power of SNP calling. Our results indicate that most sex linked regions are in contigs shared between sex chromosomes, suggesting a homomorphic architecture of *Triops* sex chromosomes. In addition, we found differences between populations with different sexual systems, such as a larger sex linked region in the gonochoric population and different Z chromosome lineages present between populations. For the first time, we identified a region of the putative sex chromosome which has undergone recent or historical recombination in the androdioecious population.

Paired end RAD sequencing utilisation

The use of second, or reverse, end RAD sequence data for genotype calling can be problematic due to its low coverage depths leading to a low confidence in calling accuracy (Davey et al., 2012). This study involved RAD data with very high coverage, along with a reference genome, which allowed us to utilise the second end sequence data whilst still using stringent filters to ensure a high confidence level in genotype calling. This approach provided an average of 7x increase in SNPs called. Second end reads can substantially increase the power of RAD sequencing, with a significantly larger region of the genome being available for analysis, even when filtered to only include the SNPs which are present in most individuals.

Our analysis, using two draft genome assemblies identified a higher number of sex linked RAD tags than a previous study in *T. cancriformis* (Mathers et al., 2015). Second end RAD reads cover a larger breadth of the genome, up to 700 bp (Davey et al., 2012), enabling us to detect sex linked SNPs further downstream from the

restriction site that would be absent in studies which only utilised 70 bp first end reads. This led to the robust characterisation of sex linked RAD tags that population analysis in Mathers et al. (2015) suggested as sex linked. The increase of confidence in genotype calling also allowed us to use depth ratios between sexes to identify Z linked RAD tags, which could not be performed *do novo* by Mathers et al. (2015). Although the number of sex linked RAD tags identified was larger, we observed fewer sex specific RAD tags in our data. One reason for this could be that Mathers et al. (2015) used Stacks v0.99993 (Catchen et al., 2011) to call genotypes from the filtered RAD data set, a software that produces artifacts when there are indels between alleles, assigning indels as different RAD tags rather than merging. To overcome this issue, we used FreeBayes (Garrison and Marth, 2012) to call SNPs, which takes into account multiple possible alignments in alleles and so can identify alleles differing by indels. This merging of indel alleles has contributed to the discrepancy between our results and those of Mathers et al. (2015). Another issue that can arise when mapping to the reference genomes is the presence of divergent gametologs or alleles between populations. If a read is sufficiently divergent from the reference assembly it was being mapped to it would be discarded from further downstream analysis. In order to avoid losing diverging W and Z specific reads, we utilised both ZZ (male) and WW (hermaphrodite) assemblies. However, data from our haplotype analysis suggests that some regions of the Z chromosome in Königswartha are sufficiently different from the Z chromosome from our Espolla male reference assembly, which might explain why identified fewer Z specific RAD tags in Königswartha than Mathers et al. (2015). In addition, we found that many of RAD tags identified by Mathers et al. (2015) as W specific were in fact artifacts of the RAD approach, as they were due to polymorphisms in the restriction site generating null Z alleles. Overall, our approach resulted in a more accurate analysis of sex specific regions of *Triops* sex chromosomes.

Sex chromosome structure in Triops cancriformis

The putative sex chromosomal regions in *T.cancriformis* behaved differently in populations with different sexual systems. We record a larger non-recombination region between Espolla sex chromosomes in agreement with Mathers et al. (2015). This reduction in the size of the non-recombining region was expected, given the predicted reduction in potential for sex antagonistic selection in Königswartha

(Mathers et al., 2015). In the androdioecious population, which has most likely transitioned from an ancestral dioecious (Zierold et al., 2007) the Z and W chromosomes spends more time in hermaphrodites than males, as the males are rare in the population (Zierold et al., 2007) and the Z is present in amphigenic hermaphrodites and the W chromosome is only present in hermaphrodite alleles, promoting female and male traits will be selected for in both chromosomes, reducing antagonistic selection (Mathers et al., 2015). In contrast, in the dioecious population with a 50:50 sex ratio, W chromosomes are only found in females and Z chromosomes spend ⅔ of their time in males, increasing the potential for sexual antagonisitic selection and increasing selection for recombination suppression.

The use of reference genomes in our study enabled, for the first time, the thorough analysis the sex specific RAD tags and determine whether they were produced through a divergent region in the chromosome or through null allele dropout generated by polymorphisms in restriction sites of RAD sequences in one particular sex (Andrews et al., 2016). Our data showed that most W specific (9/16) and some Z specific RAD tags (5/11) had been generated by null allele dropout. Despite excluding these from the estimated sex linked non-recombination region, the non-recombining region is largely constructed of contigs that are sex linked rather than sex specific, indicating that the gametologs are similar genetically and the sex chromosomes are most likely homomorphic. The homomorphic nature of sex chromosomes was suggested through a karyological study *T.cancriformis*, but the small size of the chromosomes hindered the confidence of the study (Ombretta et al., 2012). From our assembly sizes, we estimate the genome size to be 135.28 Mb (in Espolla), the largest non-recombining region we estimate is that of the W chromosome of Espolla encompassing 1.40 Mb. Karyotype analysis showed that *Triops cancriformis* possess 2n = 12 chromosomes (Ombretta et al., 2005), therefore, assuming a similar size for each chromosome, ca. 11.27 Mb, the non-recombining region would be approximately 12.42% of this. Using the higher boundary this rises to 15.44%, still suggesting a relatively small non-recombining region of the genome in comparison to the whole chromosome suggested to be under recombination suppression in another branchiopod, *Eulimnadia texana* (Weeks et al., 2010).

Sex determination systems and their genetic control can differ dramatically between

species ranging from single SNP controlled determination to large sections of the whole chromosome (Wright et al., 2017). Our results suggest that either the master sex determination gene would be a gene found in within the one sex specific contig conserved in both populations (*flattened_line_892*) or the master sex determination of *T. cancriformis* is to be controlled by a region not highly differentiated between sexes. The most likely candidate, *Sex lethal*, is found on a region of the genome that is sex linked in both populations. As we find no genes in our only W specific contig the latter is more likely and from this we can predict that sex determination is controlled by a single locus with a dominant allele determining females and hermaphrodites, and a recessive allele determining maleness, as was hypothesized for *T. newberryi* (Sassaman and Weeks, 1993; Sassaman 1995). An alternative to recombination suppression, is sex-biased expression as a resolution to sexual antagonism as is seen in some Ratites (Vicoso et al., 2012). The presence of sex-biased expression would resolve the conflict at certain loci and enable the species to adapt to the change in sexual system and the selection pressures that come with it, whilst retaining the morphologically similar chromosomes. However, in a system with sex biased expression system we would expect a reduced selection for recombination suppression across all populations, which we do not observe in the Espolla *T. cancriformis* population where we retain distinct haplotypes over a relatively large genomic region.

Evolution of sex chromosomes in Triops cancriformis

Our results supported the presence of two differentiated Z chromosomes across the populations. Mathers et al. (2015) proposed that given that only a small number of putative Z chromosomal RAD tags were shared across the populations, the divergent Z chromosome in Königswartha could have resulted from introgression from an expanding population which had differentiated in a different glacial refugia. Our results support this hypothesis, as the haplotypes of Z chromosomes from different populations are highly differentiated from one another, with the Z chromosome from Königswartha more closely resembling the W chromosome from Espolla than that of the Z from Espolla. Mathers et al. (2015) showed that the genotypes of the RAD tags sex linked in both populations showed patterns suggesting putative W alleles were similar across populations. Our results revealed a slightly different pattern, that there

are in fact two putative W haplotypes present across the populations, one found in Espolla and Königswartha and another only found in a number of individuals in Königswartha. We propose that this additional W haplotype, found only in Königswartha, has originated through either a historical recombination event between a region of the W chromosome and a region of the Z chromosome in Königswartha or that there is an additional introgression from another population, or possibly a combination of both. Comparatively, this W haplotype found only in Königswartha sits between the W haplotype found in Espolla and the Z found only in Königswartha, a mtDNA study by Zierold et al. (2007) detailed that the lineages of Espolla and Königswartha are shared, with the Espolla-containing Iberian Peninsula populations most likely being a refuge during the Pleistocene and a recent evolution of androdioecy and hermaphroditism has aided the European expansion. However, without mtDNA for the individuals for which we sequenced and analysis using outgroups and further populations, we are unable to clarify the full evolutionary history of the populations.

Sexual antagonistic selection is hypothesized to be the prevalent force in driving of evolution of sex chromosomes, increasing selection for recombination suppression to enhance the canalisation of sexes (Beukeboom & Perrin, 2014). However, Muller's ratchet can lead to an accumulation of deleterious mutations on non-recombining regions (Charlesworth & Charlesworth, 2000). Different organisms have evolved methods to reduce the impact of these deleterious mutations in homomorphic chromosomes. For instance, in frogs, sex reversed individuals undergo rare recombination events in homomorphic frogs. Frogs display extreme heterochiasmia, a phenomenon by which different levels of recombination occur in different sexes, with females undergoing much more recombination than males. Heterochiasmia, however, is determined by the phenotypic, not genotypic sex, sex reversed XY females undergo increased recombination between X and Y chromosomes, which in evolutionary timescales is thought to preserve chromosome homomorphism (Perrin, 2009). In addition, fluctuating sex ratios have stimulated the increase of recombination in the pseudoautosomal regions to remove deleterious mutations in homomorphic chromosomes of a wild strawberry (Tennessen et al., 2016), with more biased sex ratios leading to increased recombination to negate the effects of a low effective population size. Anticipating that the contigs that have been affected by this

recombination event in *T. cancriformis* are physically linked we expect to see the emergence of evolutionary strata in the sex chromosomes of *T.cancriformis* within Königswartha. Evolutionary strata occur after a recombination arrest event between the sex chromosomes. Through analysis of the strata details of the time and processes involved in the recombination suppression can be uncovered (Lawson-Handley et al., 2003). Our results show strata within the W chromosome only present in Königswartha, with one strata sex specific, conserved and sex linked in both populations, the next strata is non-recombining but showing evidence of a historical recombination event, the third is non-sex linked but still providing evidence of a past recombination event, and the last strata is not sex linked and conserved throughout W and Z chromosomes (Figure 11).

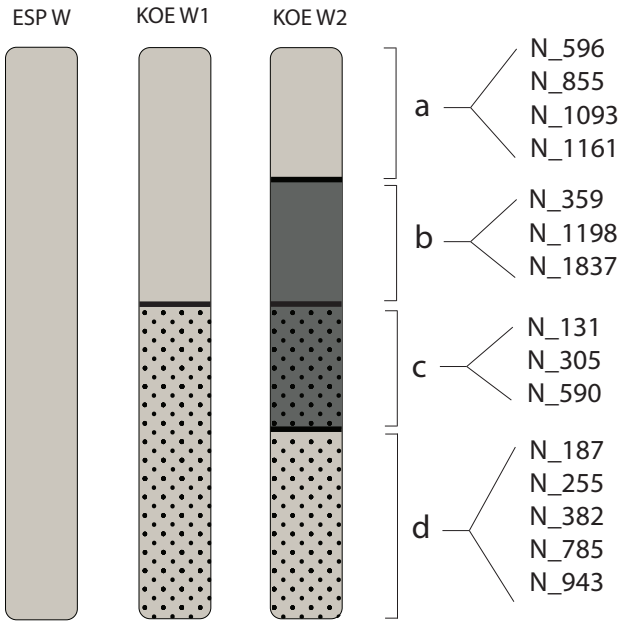


Figure 11. An illustration of different W chromosomes found in the study for visual purposes, size of 'strata' is not accurate. The 'strata' consist of a conserved region across populations (a); a differentiated region in Königswartha separate from that in Espolla, part is under recombination suppression (b) and part is not (c); a region of the chromosome is conserved between populations but not under recombination suppression in Königswartha (d). Dotted region represents area not under recombination suppression.

Evolution of androdioecy

Our results note disparities in genomic composition between both populations that might reflect the two sexual systems, but might also be affected by population history (e.g. phylogeography, bottlenecks). We see a much larger genomic polymorphism in the dioecious Espolla than the androdioecious Königswartha population with SNPs under full linkage over greater genetic distances in the androdioecious population. The high linkage disequilibrium and low polymorphisms over the genome of the androdioecious population is in line with what we would anticipate from a selfing population, with low effective recombination rates and low levels of variation between individuals (Nordborg, 1997) and outcrossing might be expected to be selected for to increase selective recombination. Although this pattern can also arise from population bottlenecks, with the low effective recombination and low polymorphism frequency being a product (Shaper et al., 2012). These results highlight consequences of selfing populations but also the supports the hypothesis of a recent colonisation of a northward spread of populations inline with glacial retreat (Zierold et al., 2007). Our results represent genetic patterns that have been seen in many other biota having experienced postglacial recolonisation (Hewitt, 2004; 2008).

Limitations

One of the assumptions to our study is that the sex linked regions within *T. cancriformis* are physically linked on a single sex chromosome, as for our analysis we have worked with fragmented reference genome assembly brought about by repetitive regions and as a result we cannot analyse the physical structure of the proposed single sex chromosome. To test this hypothesis we would need to use PCR targeted enrichment and long-range sequencing (Uribe-Confers et al., 2016), targeting the contigs discovered in this study, this would also help clarify the structure of the different stata. In addition, the reference assemblies in this study would have also been advantageous if they had been generated in the same way, mapping to similar contigs and minimising the effect of the assembly process in clarifying the homomorphism of the sex linked regions. This would have also resolved the necessary step of the production of a dictionary to cross reference contigs.

Future direction

To further the understanding of *Triops cancriformis* sex chromosomes it would be advantageous to relax the criteria for mapping to assemblies, by allowing multiple hits on different regions of the genome to be be able to identify duplicated scaffolds and bioinformatic techniques could used *a posteriori* to remove non sex-linked repetitive regions. In reference to the different haplotypes present within Königswartha it would be beneficial to target sex linked contigs of individuals representative of the different haplotypes, enrich and sequence these regions, clarifying the differences within the population. Long reads would be beneficial to analyse the structure and composition of the regions around the contigs so it would be useful to use an ultra-long read sequencer e.g. Nanopore Minion (Jain et al., 2016) or optical mapping (Dong et al., 2013). In our study we identified a panel of different genes that could contribute to sex determination in *Triops cancriformis* to further analyse this it would be useful to use transcriptomics. We have identified contigs that could be used to synthesise sex specific primers that could be utilised to identify sex during the developing diapause stage, performing RNA analysis with the annotated genome to determine sex biased expression rates could assist in determining the locus or gene responsible for sex determination in both populations.

Conclusions

Due to the differing sexual systems, *Triops cancriformis* offer a novel opportunity to investigate the impacts on the architecture of sex chromosomes caused by sexually antagonistic selection and changes in recombination levels. Our study incorporated two reference genome assemblies allowing 2nd end RAD data to be used to analyse the genome increasing the efficiency and the scope of RAD data that has been used in a previous study (Mathers et al., 2015). This allowed us to more accurately analyse the sex linked markers present in *T. cancriformis* and our data has suggested that the sex chromosomes are homomorphic. Our study has uncovered details of the expansion of populations, with Z chromosomes in different populations are likely from differing origins, despite a W chromosome shared between populations. And that a recombination event occurred in a sex linked region in the hermaphroditic population,

although an outgroup would be needed to confirm this. In addition, we have identified a number of genes that could be involved in the sex determination pathway but transcriptomic analysis is required to identify the gene, or group of genes, responsible for sex determination and or involved in the transition to hermaphroditism and affected by changing patterns of sexual antagonistic selection.

Bibliography

AITKEN, R.J. and MARSHALL GRAVES, J.A., 2002. Human spermatozoa: The future of sex. Nature. 415, p. 963.

ALTSCHUL, S.F., GISH, W., MILLER, W., MYERS, E.W., and LIPMAN, D.J., 1990. Basic local alignment search tool. Journal of molecular biology. 215 (3), pp. 403–410.

ANDERSON, J.L., MARÍ, A.R., BRAASCH, I., AMORES, A., HOHENLOHE, P., BATZEL, P., and POSTLETHWAIT, J.H., 2012. Multiple Sex-Associated Regions and a Putative Sex Chromosome in Zebrafish Revealed by RAD Mapping and Population Genomics. PloS one. 7 (7), p. e40701.

ANDREWS, K.R., GOOD, J.M., MILLER, M.R., LUIKART, G., and HOHENLOHE, P.A., 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nature reviews. Genetics. 17 (2), pp. 81–92.

ANTIPOV, D., KOROBEYNIKOV, A., MCLEAN, J.S., and PEVZNER, P.A., 2016. hybridSPAdes: an algorithm for hybrid assembly of short and long reads. Bioinformatics . 32 (7), pp. 1009–1015.

ARNQVIST, G. and ROWE, L., 2005. Sexual Conflict: Princeton University Press.

AVISE, J., 2011. Hermaphroditism: A Primer on the Biology, Ecology, and Evolution of Dual Sexuality. Columbia University Press.

BACHTROG, D. and CHARLESWORTH, B., 2002. Reduced adaptation of a non-recombining neo-Y chromosome. Nature. 416 (6878), pp. 323–326.

BACHTROG, D., KIRKPATRICK, M., MANK, J.E., MCDANIEL, S.F., PIRES, J.C., RICE, W., and VALENZUELA, N., 2011. Are all sex chromosomes created equal? Trends in genetics: TIG. 27 (9), pp. 350–357.

BACHTROG, D., MANK, J.E., PEICHEL, C.L., KIRKPATRICK, M., OTTO, S.P., ASHMAN, T.-L., HAHN, M.W., KITANO, J., MAYROSE, I., MING, R., PERRIN, N., ROSS, L., VALENZUELA, N., VAMOSI, J.C., and TREE OF SEX CONSORTIUM, 2014. Sex determination: why so many ways of doing it? PLoS biology. 12 (7), p. e1001899.

BAIRD, N.A., ETTER, P.D., ATWOOD, T.S., CURREY, M.C., SHIVER, A.L., LEWIS, Z.A., SELKER, E.U., CRESKO, W.A., and JOHNSON, E.A., 2008. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. PloS one. 3 (10), p. e3376.

BAXTER, S.W., DAVEY, J.W., SPENCER JOHNSTON, J., SHELTON, A.M., HECKEL, D.G., JIGGINS, C.D., and BLAXTER, M.L., 2011. Linkage Mapping and Comparative Genomics Using Next-Generation RAD Sequencing of a Non-Model

Organism. PloS one. 6 (4), p. e19315.

BEISSINGER, T.M., HIRSCH, C.N., SEKHON, R.S., FOERSTER, J.M., JOHNSON, J.M., MUTTONI, G., VAILLANCOURT, B., ROBIN BUELL, C., KAEPPLER, S.M., and DE LEON, N., 2013. Marker Density and Read Depth for Genotyping Populations Using Genotyping-by-Sequencing. Genetics. 193 (4), p. 1073.

BELLOTT, D.W., HUGHES, J.F., SKALETSKY, H., BROWN, L.G., PYNTIKOVA, T., CHO, T.-J., KOUTSEVA, N., ZAGHLUL, S., GRAVES, T., ROCK, S., KREMITZKI, C., FULTON, R.S., DUGAN, S., DING, Y., MORTON, D., KHAN, Z., LEWIS, L., BUHAY, C., WANG, Q., WATT, J., HOLDER, M., LEE, S., NAZARETH, L., ALFÖLDI, J., ROZEN, S., MUZNY, D.M., WARREN, W.C., GIBBS, R.A., WILSON, R.K., and PAGE, D.C., 2014. Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. Nature. 508 (7497), pp. 494–499.

BENSON, D.A., KARSCH-MIZRACHI, I., LIPMAN, D.J., OSTELL, J., and WHEELER, D.L., 2005. GenBank. Nucleic acids research. 33 (Database issue), pp. D34–8.

BÖHNE, A., WILSON, C.A., POSTLETHWAIT, J.H., and SALZBURGER, W., 2016. Variations on a theme: Genomics of sex determination in the cichlid fish Astatotilapia burtoni. BMC genomics. 17 (1), p. 883.

BRELSFORD, A., LAVANCHY, G., SERMIER, R., RAUSCH, A., and PERRIN, N., 2016. Identifying homomorphic sex chromosomes from wild-caught adults with limited genomic resources. Molecular ecology resources [online]. Available from: http://dx.doi.org/10.1111/1755-0998.12624.

BULL, J.J., 1983. Evolution of sex determining mechanisms. Benjamin/Cummings Pub. Co.

CATCHEN, J., HOHENLOHE, P.A., BASSHAM, S., AMORES, A., and CRESKO, W.A., 2013. Stacks: an analysis tool set for population genomics. Molecular ecology. 22 (11), pp. 3124–3140.

CHARLESWORTH, B. and CHARLESWORTH, D., 1978. A Model for the Evolution of Dioecy and Gynodioecy. The American naturalist. 112 (988), pp. 975–997.

CHARLESWORTH, B. and CHARLESWORTH, D. (2000) ‘The degeneration of Y chromosomes’, Philosophical Transactions of the Royal Society B: Biological Sciences, 355(1403), pp. 1563–1572.

CHARLESWORTH, D. and WRIGHT, S.I., 2001. Breeding systems and genome evolution. Current opinion in genetics & development. 11 (6), pp. 685–690.

CHASNOV, J.R., 2010. The evolution from females to hermaphrodites results in a sexual conflict over mating in androdioecious nematode worms and clam shrimp. Journal of evolutionary biology. 23 (3), pp. 539–556.

CHONG, Z., RUAN, J., and WU, C.-I., 2012. Rainbow: an integrated tool for efficient clustering and assembling RAD-seq reads. Bioinformatics . 28 (21), pp. 2732–2737.

DANECEK, P., AUTON, A., ABECASIS, G., ALBERS, C.A., BANKS, E., DEPRISTO, M.A., HANDSAKER, R.E., LUNTER, G., MARTH, G.T., SHERRY, S.T., MCVEAN, G., DURBIN, R., and 1000 GENOMES PROJECT ANALYSIS GROUP, 2011. The variant call format and VCFtools. Bioinformatics . 27 (15), pp. 2156–2158.

DAVEY, J.W., CEZARD, T., FUENTES-UTRILLA, P., ELAND, C., GHARBI, K., and BLAXTER, M.L., 2013. Special features of RAD Sequencing data: implications for genotyping. Molecular ecology. 22 (11), pp. 3151–3164.

BEUKEBOOM, L.W. and PERRIN, N. (2014) 'What are sexes, and why are there sexes?', in The Evolution of Sex Determination. Oxford University Press (OUP), pp. 1–17.

DAVEY, J.W., DAVEY, J.L., BLAXTER, M.L., and BLAXTER, M.W., 2010. RADSeq: next-generation population genetics. Briefings in functional genomics. 9 (5-6), pp. 416–423.

DELANEAU, O., HOWIE, B., COX, A.J., ZAGURY, J.-F., and MARCHINI, J., 2013. Haplotype estimation using sequencing reads. American journal of human genetics. 93 (4), pp. 687–696.

DEVLIN, R.H. and NAGAHAMA, Y., 2002. Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. Aquaculture . 208 (3), pp. 191–364.

DONG, Y., XIE, M., JIANG, Y., XIAO, N., DU, X., ZHANG, W., TOSSER-KLOPP, G., WANG, J., YANG, S., LIANG, J., CHEN, W., CHEN, J., ZENG, P., HOU, Y., BIAN, C., PAN, S., LI, Y., LIU, X., WANG, W., SERVIN, B., SAYRE, B., ZHU, B., SWEENEY, D., MOORE, R., NIE, W., SHEN, Y., ZHAO, R., ZHANG, G., LI, J., FARAUT, T., WOMACK, J., ZHANG, Y., KIJAS, J., COCKETT, N., XU, X., ZHAO, S., WANG, J., and WANG, W., 2013. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (Capra hircus). Nature biotechnology. 31 (2), pp. 135–141.

EMERSON, K.J., MERZ, C.R., CATCHEN, J.M., HOHENLOHE, P.A., CRESKO, W.A., BRADSHAW, W.E., and HOLZAPFEL, C.M., 2010. Resolving postglacial phylogeography using high-throughput sequencing. Proceedings of the National Academy of Sciences. 107 (37), pp. 16196–16200.

FOSTER, J.W. and GRAVES, J.A., 1994. An SRY-related sequence on the marsupial X chromosome: implications for the evolution of the mammalian testis-determining gene. Proceedings of the National Academy of Sciences of the United States of America. 91 (5), pp. 1927–1931.

G. FRYER, F.R.S., 1988. Studies on the functional morphology and biology of the Notostraca (Crustacea: Branchiopoda). Philosophical transactions of the Royal Society of London. Series B, Biological sciences. 321 (1203), pp. 27–124.

GARRISON, E. and MARTH, G., 2012. Haplotype-based variant detection from short-read sequencing. arXiv [q-bio.GN] [online]. Available from: http://arxiv.org/abs/1207.3907.

GHISELIN, M.T., 1969. The evolution of hermaphroditism among animals. The Quarterly review of biology. 44 (2), pp. 189–208.

GÖTZ, S., GARCÍA-GÓMEZ, J.M., TEROL, J., WILLIAMS, T.D., NAGARAJ, S.H., NUEDA, M.J., ROBLES, M., TALÓN, M., DOPAZO, J., and CONESA, A., 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic acids research. 36 (10), pp. 3420–3435.

HANDLEY, L.-J.L., CEPLITIS, H., and ELLEGREN, H., 2004. Evolutionary strata on the chicken Z chromosome: implications for sex chromosome evolution. Genetics. 167 (1), pp. 367–376.

HILL, W.G. and ROBERTSON, A., 1966. The effect of linkage on limits to artificial selection. Genetical research. 8 (3), pp. 269–294.

HOHENLOHE, P.A., PHILLIPS, P.C., and CRESKO, W.A., 2010. USING POPULATION GENOMICS TO DETECT SELECTION IN NATURAL POPULATIONS: KEY CONCEPTS AND METHODOLOGICAL CONSIDERATIONS. International journal of plant sciences. 171 (9), pp. 1059–1071.

HOLT, C. and YANDELL, M., 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC bioinformatics. 12, p. 491.

HORN, R.L., RAMARAJ, T., DEVITT, N.P., SCHILKEY, F.D., and COWLEY, D.E., 2017. De novo assembly of a tadpole shrimp (Triops newberryi) transcriptome and preliminary differential gene expression analysis. Molecular ecology resources. 17 (2), pp. 161–171.

JAIN, M., OLSEN, H.E., PATEN, B., and AKESON, M., 2016. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome biology. 17 (1), p. 239.

JANZEN, F.J. and PAUKSTIS, G.L., 1991. Environmental Sex Determination in Reptiles: Ecology, Evolution, and Experimental Design. The Quarterly review of biology. 66 (2), pp. 149–179.

JARNE, P. and AULD, J.R., 2006. Animals mix it up too: the distribution of self-fertilization among hermaphroditic animals. Evolution; international journal of organic evolution. 60 (9), pp. 1816–1824.

JONES, P., BINNS, D., CHANG, H.-Y., FRASER, M., LI, W., MCANULLA, C., MCWILLIAM, H., MASLEN, J., MITCHELL, A., NUKA, G., PESSEAT, S., QUINN, A.F., SANGRADOR-VEGAS, A., SCHEREMETJEW, M., YONG, S.-Y., LOPEZ, R., and HUNTER, S., 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics . 30 (9), pp. 1236–1240.

JORDAN, C.Y. and CONNALLON, T., 2014. Sexually antagonistic polymorphism in simultaneous hermaphrodites. Evolution; international journal of organic evolution. 68 (12), pp. 3555–3569.

KORF, I., 2004. Gene finding in novel genomes. BMC bioinformatics. 5, p. 59.

KORN, M., MARRONE, F., PÉREZ-BOTE, J.L., MACHADO, M., CRISTO, M., DA FONSECA, L.C., and HUNDSDOERFER, A.K., 2006. Sister species within the Triops cancriformis lineage (Crustacea, Notostraca). Zoologica scripta. 35 (4), pp. 301–322.

KORN, M., RABET, N., GHATE, H.V., MARRONE, F., and HUNDSDOERFER, A.K., 2013. Molecular phylogeny of the Notostraca. Molecular phylogenetics and evolution. 69 (3), pp. 1159–1171.

LARSSON, A., 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics . 30 (22), pp. 3276–3278.

LEE, P.L.M., BRAIN, P.F., FORMAN, D.W., BRADBURY, R.B., and GRIFFITHS, R., 2002. Sex and death: CHD1Z associated with high mortality in moorhens. Evolution; international journal of organic evolution. 56 (12), pp. 2548–2553.

LI, H. and DURBIN, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics . 25 (14), pp. 1754–1760.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R., and 1000 GENOME PROJECT DATA PROCESSING SUBGROUP, 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics . 25 (16), pp. 2078–2079.

LONGHURST, A.R., 1955. A review of the Notostraca. Bulletin of the British Museum of Natural History, Botany. 3, pp. 1–57.

LOVE, R.R., WEISENFELD, N.I., JAFFE, D.B., BESANSKY, N.J., and NEAFSEY, D.E., 2016. Evaluation of DISCOVAR de novo using a mosquito sample for cost-effective short-read genome assembly. BMC genomics. 17, p. 187.

MATHERS, T.C., HAMMOND, R.L., JENNER, R.A., HÄNFLING, B., ATKINS, J., and GÓMEZ, A., 2015. Transition in sexual system and sex chromosome evolution in the tadpole shrimp Triops cancriformis. Heredity. 115 (1), pp. 37–46.

MATHERS, T.C., HAMMOND, R.L., JENNER, R.A., HÄNFLING, B., and GÓMEZ, A., 2013. Multiple global radiations in tadpole shrimps challenge the concept of ‘living fossils’. PeerJ. 1, p. e62.

MATHERS, T.C., HAMMOND, R.L., JENNER, R.A., ZIEROLD, T., HÄNFLING, B., and GÓMEZ, A., 2013. High lability of sexual system over 250 million years of evolution in morphologically conservative tadpole shrimps. BMC evolutionary biology. 13 (1), p. 30.

MEUSEL, F. and SCHWENTNER, M., 2016. Molecular and morphological delimitation of Australian Triops species (Crustacea: Branchiopoda: Notostraca)—large diversity and little morphological differentiation. Organisms, diversity & evolution. 1 (17), pp. 137–156.

MITCHELL, N.J. and JANZEN, F.J., 2010. Temperature-dependent sex determination and contemporary climate change. Sexual development: genetics, molecular biology, evolution, endocrinology, embryology, and pathology of sex determination and differentiation. 4 (1-2), pp. 129–140.

MYOSHO, T., OTAKE, H., MASUYAMA, H., MATSUDA, M., KUROKI, Y., FUJIYAMA, A., NARUSE, K., HAMAGUCHI, S., and SAKAIZUMI, M., 2012. Tracing the emergence of a novel sex-determining gene in medaka, Oryzias luzonensis. Genetics. 191 (1), pp. 163–170.

NORDBORG, M., 1997. Structured coalescent processes on different time scales. Genetics. 146 (4), pp. 1501–1514.

NURK, S., BANKEVICH, A., ANTIPOV, D., GUREVICH, A., KOROBEYNIKOV, A., LAPIDUS, A., PRJIBELSKY, A., PYSHKIN, A., SIROTKIN, A., SIROTKIN, Y., STEPANAUSKAS, R., MCLEAN, J., LASKEN, R., CLINGENPEEL, S.R., WOYKE, T., TESLER, G., ALEKSEYEV, M.A., and PEVZNER, P.A., 2013. Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads. In: Research in Computational Molecular Biology. Presented at the Annual International Conference on Research in Computational Molecular Biology, Springer, Berlin, Heidelberg. pp. 158–170.

OGDEN, R., GHARBI, K., MUGUE, N., MARTINSOHN, J., SENN, H., DAVEY, J.W., POURKAZEMI, M., MCEWING, R., ELAND, C., VIDOTTO, M., SERGEEV, A., and CONGIU, L., 2013. Sturgeon conservation genomics: SNP discovery and validation using RAD sequencing. Molecular ecology. 22 (11), pp. 3112–3123.

OMBRETTA, M., CESARI, M., EDER, E., SCANABISSI, F., and MANTOVANI, B., 2005. Chromosomes in sexual populations of Notostracan and Con-chostracan taxa (Crustacea, Branchiopoda). Caryologia. 58 (2), pp. 164–170.

OSTBERG, C.O., HAUSER, L., PRITCHARD, V.L., GARZA, J.C., and NAISH, K.A., 2013. Chromosome rearrangements, recombination suppression, and limited segregation distortion in hybrids between Yellowstone cutthroat trout (Oncorhynchus clarkii bouvieri) and rainbow trout (O. mykiss). BMC genomics. 14, p. 570.

PAN, Z.-J., LI, X.-Y., ZHOU, F.-J., QIANG, X.-G., and GUI, J.-F., 2015. Identification of Sex-Specific Markers Reveals Male Heterogametic Sex Determination in Pseudobagrus ussuriensis. Marine biotechnology . 17 (4), pp. 441–451.

PANNELL, J.R., 2002. The Evolution and Maintenance of Androdioecy. Annual review of ecology and systematics. 33 (1), pp. 397–425.

PANNELL, J.R. and BARRETT, S.C.H., 1998. BAKER'S LAW REVISITED:

REPRODUCTIVE ASSURANCE IN A METAPOPULATION. Evolution; international journal of organic evolution. 52 (3), pp. 657–668.

PARRA, G., BRADNAM, K., and KORF, I., 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics . 23 (9), pp. 1061–1067.

PERRIN, N., 2009. Sex reversal: a fountain of youth for sex chromosomes? Evolution; international journal of organic evolution. 63 (12), pp. 3043–3049.

PURCELL, S., NEALE, B., TODD-BROWN, K., THOMAS, L., FERREIRA, M.A.R., BENDER, D., MALLER, J., SKLAR, P., DE BAKKER, P.I.W., DALY, M.J., and SHAM, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics. 81 (3), pp. 559–575.

R CORE TEAM, 2016. A language and environment for statistical computing. R Foundation for statistical computing, 2015; Vienna, Austria.

REIMAND, J., KULL, M., PETERSON, H., HANSEN, J., and VILO, J., 2007. g:Profiler--a web-based toolset for functional profiling of gene lists from large-scale experiments. Nucleic acids research. 35 (Web Server issue), pp. W193–200.

RICE, W.R., 1987. THE ACCUMULATION OF SEXUALLY ANTAGONISTIC GENES AS A SELECTIVE AGENT PROMOTING THE EVOLUTION OF REDUCED RECOMBINATION BETWEEN PRIMITIVE SEX CHROMOSOMES. Evolution; international journal of organic evolution. 41 (4), pp. 911–914.

RICE, W.R., 1996. Evolution of the Y Sex Chromosome in Animals. Bioscience. 46 (5), pp. 331–343.

ROSS, M.T., GRAFHAM, D.V., COFFEY, A.J., SCHERER, S., MCLAY, K., MUZNY, D., PLATZER, M., HOWELL, G.R., BURROWS, C., BIRD, C.P., FRANKISH, A., LOVELL, F.L., HOWE, K.L., ASHURST, J.L., FULTON, R.S., SUDBRAK, R., WEN, G., JONES, M.C., HURLES, M.E., ANDREWS, T.D., SCOTT, C.E., SEARLE, S., RAMSER, J., WHITTAKER, A., DEADMAN, R., CARTER, N.P., HUNT, S.E., CHEN, R., CREE, A., GUNARATNE, P., HAVLAK, P., HODGSON, A., METZKER, M.L., RICHARDS, S., SCOTT, G., STEFFEN, D., SODERGREN, E., WHEELER, D.A., WORLEY, K.C., AINSCOUGH, R., AMBROSE, K.D., ANSARI-LARI, M.A., ARADHYA, S., ASHWELL, R.I.S., BABBAGE, A.K., BAGGULEY, C.L., BALLABIO, A., BANERJEE, R., BARKER, G.E., BARLOW, K.F., BARRETT, I.P., BATES, K.N., BEARE, D.M., BEASLEY, H., BEASLEY, O., BECK, A., BETHEL, G., BLECHSCHMIDT, K., BRADY, N., BRAY-ALLEN, S., BRIDGEMAN, A.M., BROWN, A.J., BROWN, M.J., BONNIN, D., BRUFORD, E.A., BUHAY, C., BURCH, P., BURFORD, D., BURGESS, J., BURRILL, W., BURTON, J., BYE, J.M., CARDER, C., CARREL, L., CHAKO, J., CHAPMAN, J.C., CHAVEZ, D., CHEN, E., CHEN, G., CHEN, Y., CHEN, Z., CHINAULT, C., CICCODICOLA, A., CLARK, S.Y., CLARKE, G., CLEE, C.M., CLEGG, S., CLERC-BLANKENBURG, K., CLIFFORD, K., COBLEY, V., COLE, C.G., CONQUER, J.S., CORBY, N., CONNOR, R.E., DAVID, R., DAVIES, J., DAVIS, C., DAVIS, J., DELGADO, O., DESHAZO, D., DHAMI, P., DING, Y., DINH, H., DODSWORTH, S., DRAPER, H., DUGAN-ROCHA, S., DUNHAM, A., DUNN, M., DURBIN, K.J., DUTTA, I., EADES, T., ELLWOOD, M.,

EMERY-COHEN, A., ERRINGTON, H., EVANS, K.L., FAULKNER, L., FRANCIS, F., FRANKLAND, J., FRASER, A.E., GALGOCZY, P., GILBERT, J., GILL, R., GLÖCKNER, G., GREGORY, S.G., GRIBBLE, S., GRIFFITHS, C., GROCOCK, R., GU, Y., GWILLIAM, R., HAMILTON, C., HART, E.A., HAWES, A., HEATH, P.D., HEITMANN, K., HENNIG, S., HERNANDEZ, J., HINZMANN, B., HO, S., HOFFS, M., HOWDEN, P.J., HUCKLE, E.J., HUME, J., HUNT, P.J., HUNT, A.R., ISHERWOOD, J., JACOB, L., JOHNSON, D., JONES, S., DE JONG, P.J., JOSEPH, S.S., KEENAN, S., KELLY, S., KERSHAW, J.K., KHAN, Z., KIOSCHIS, P., KLAGES, S., KNIGHTS, A.J., KOSIURA, A., KOVAR-SMITH, C., LAIRD, G.K., LANGFORD, C., LAWLOR, S., LEVERSHA, M., LEWIS, L., LIU, W., LLOYD, C., LLOYD, D.M., LOULSEGED, H., LOVELAND, J.E., LOVELL, J.D., LOZADO, R., LU, J., LYNE, R., MA, J., MAHESHWARI, M., MATTHEWS, L.H., MCDOWALL, J., MCLAREN, S., MCMURRAY, A., MEIDL, P., MEITINGER, T., MILNE, S., MINER, G., MISTRY, S.L., MORGAN, M., MORRIS, S., MÜLLER, I., MULLIKIN, J.C., NGUYEN, N., NORDSIEK, G., NYAKATURA, G., O'DELL, C.N., OKWUONU, G., PALMER, S., PANDIAN, R., PARKER, D., PARRISH, J., PASTERNAK, S., PATEL, D., PEARCE, A.V., PEARSON, D.M., PELAN, S.E., PEREZ, L., PORTER, K.M., RAMSEY, Y., REICHWALD, K., RHODES, S., RIDLER, K.A., SCHLESSINGER, D., SCHUELER, M.G., SEHRA, H.K., SHAW-SMITH, C., SHEN, H., SHERIDAN, E.M., SHOWNKEEN, R., SKUCE, C.D., SMITH, M.L., SOTHERAN, E.C., STEINGRUBER, H.E., STEWARD, C.A., STOREY, R., SWANN, R.M., SWARBRECK, D., TABOR, P.E., TAUDIEN, S., TAYLOR, T., TEAGUE, B., THOMAS, K., THORPE, A., TIMMS, K., TRACEY, A., TREVANION, S., TROMANS, A.C., D'URSO, M., VERDUZCO, D., VILLASANA, D., WALDRON, L., WALL, M., WANG, Q., WARREN, J., WARRY, G.L., WEI, X., WEST, A., WHITEHEAD, S.L., WHITELEY, M.N., WILKINSON, J.E., WILLEY, D.L., WILLIAMS, G., WILLIAMS, L., WILLIAMSON, A., WILLIAMSON, H., WILMING, L., WOODMANSEY, R.L., WRAY, P.W., YEN, J., ZHANG, J., ZHOU, J., ZOGHBI, H., ZORILLA, S., BUCK, D., REINHARDT, R., POUSTKA, A., ROSENTHAL, A., LEHRACH, H., MEINDL, A., MINX, P.J., HILLIER, L.W., WILLARD, H.F., WILSON, R.K., WATERSTON, R.H., RICE, C.M., VAUDIN, M., COULSON, A., NELSON, D.L., WEINSTOCK, G., SULSTON, J.E., DURBIN, R., HUBBARD, T., GIBBS, R.A., BECK, S., ROGERS, J., and BENTLEY, D.R., 2005. The DNA sequence of the human X chromosome. Nature. 434 (7031), pp. 325–337.

RUTKOWSKA, J. and BADYAEV, A.V., 2008. Review. Meiotic drive and sex determination: molecular and cytological mechanisms of sex ratio adjustment in birds. Philosophical transactions of the Royal Society of London. Series B, Biological sciences. 363 (1497), pp. 1675–1686.

SAKAI, A.K., WELLER, S.G., CHEN, M.-L., CHOU, S.-Y., and TASANONT, C., 1997. Evolution of Gynodioecy and Maintenance of Females: The Role of Inbreeding Depression, Outcrossing Rates, and Resource Allocation in Schiedea adamantis (Caryophyllaceae). Evolution; international journal of organic evolution. 51 (3), pp. 724–736.

SASSAMAN, C., 1991. Sex ratio variation in female-biased populations of Notostracans. In: Studies on Large Branchiopod Biology and Aquaculture. Springer, Dordrecht. pp. 169–179.

SASSAMAN, C., 1995. Sex determination and evolution of unisexuality in the

Conchostraca. In: Studies on Large Branchiopod Biology and Aquaculture II. Springer. pp. 45–65.

SASSAMAN, C. and WEEKS, S.C., 1993. The genetic mechanism of sex determination in the conchostracan shrimp Eulimnadia texana. The American naturalist. 141 (2), pp. 314–328.

SHAFER, A.B.A., PEART, C.R., TUSSO, S., MAAYAN, I., BRELSFORD, A., WHEAT, C.W., and WOLF, J.B.W., 2017. Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. Methods in ecology and evolution / British Ecological Society. 8 (8), pp. 907–917.

SKALETSKY, H., KURODA-KAWAGUCHI, T., MINX, P.J., CORDUM, H.S., HILLIER, L., BROWN, L.G., REPPING, S., PYNTIKOVA, T., ALI, J., BIERI, T., CHINWALLA, A., DELEHAUNTY, A., DELEHAUNTY, K., DU, H., FEWELL, G., FULTON, L., FULTON, R., GRAVES, T., HOU, S.-F., LATRIELLE, P., LEONARD, S., MARDIS, E., MAUPIN, R., MCPHERSON, J., MINER, T., NASH, W., NGUYEN, C., OZERSKY, P., PEPIN, K., ROCK, S., ROHLFING, T., SCOTT, K., SCHULTZ, B., STRONG, C., TIN-WOLLAM, A., YANG, S.-P., WATERSTON, R.H., WILSON, R.K., ROZEN, S., and PAGE, D.C., 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature. 423 (6942), pp. 825–837.

STANKE, M. and MORGENSTERN, B., 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic acids research. 33 (Web Server issue), pp. W465–7.

STEINEMANN, S. and STEINEMANN, M., 2005. Y chromosomes: born to be destroyed. BioEssays: news and reviews in molecular, cellular and developmental biology. 27 (10), pp. 1076–1083.

STÖCK, M., HORN, A., GROSSEN, C., LINDTKE, D., SERMIER, R., BETTO-COLLIARD, C., DUFRESNES, C., BONJOUR, E., DUMAS, Z., LUQUET, E., MADDALENA, T., SOUSA, H.C., MARTINEZ-SOLANO, I., and PERRIN, N., 2011. Ever-Young Sex Chromosomes in European Tree Frogs. PLoS biology. 9 (5), p. e1001062.

TENNESSEN, J.A., GOVINDARAJULU, R., LISTON, A., and ASHMAN, T.-L., 2016. Homomorphic ZW chromosomes in a wild strawberry show distinctive recombination heterogeneity but a small sex-determining region. The New phytologist. 211 (4), pp. 1412–1423.

THE UNIPROT CONSORTIUM, 2017. UniProt: the universal protein knowledgebase. Nucleic acids research. 45 (D1), pp. D158–D169.

THIÉRY, A., 1997. Horizontal distribution and abundance of cysts of several large branchiopods in temporary pool and ditch sediments. Hydrobiologia. 359 (1-3), pp. 177–189.

TRAUT, W. and MAREC, F., 1997. Sex chromosome differentiation in some species

of Lepidoptera (Insecta). Chromosome research: an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology. 5 (5), pp. 283–291.

TRUSHEIM, F. 1938. Triopsidew aus dem Keuper-Frankens. Paleont. Z., 19, pp. 198-216

ÚBEDA, F., PATTEN, M.M., and WILD, G., 2015. On the origin of sex chromosomes from meiotic drive. Proceedings. Biological sciences / The Royal Society. 282 (1798), p. 20141932.

URIBE-CONVERS, S., SETTLES, M.L., and TANK, D.C., 2016. A Phylogenomic Approach Based on PCR Target Enrichment and High Throughput Sequencing: Resolving the Diversity within the South American Species of Bartsia L. (Orobanchaceae). PloS one. 11 (2), p. e0148203.

VANSCHOENWINKEL, B., PINCEEL, T., VANHOVE, M.P.M., DENIS, C., JOCQUE, M., TIMMS, B.V., and BRENDONCK, L., 2012. Toward a Global Phylogeny of the “Living Fossil" Crustacean Order of the Notostraca. PloS one. 7 (4), p. e34998.

VICOSO, B., KAISER, V.B., and BACHTROG, D., 2013. Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution. Proceedings of the National Academy of Sciences of the United States of America. 110 (16), pp. 6453–6458.

WEEKS, S.C., 2012. The role of androdioecy and gynodioecy in mediating evolutionary transitions between dioecy and hermaphroditism in the animalia. Evolution; international journal of organic evolution. 66 (12), pp. 3670–3686.

WEEKS, S.C., BENVENUTO, C., REED, S.K., DUFF, R.J., DUAN, Z.-H., and DAVID, P., 2014. A field test of a model for the stability of androdioecy in the freshwater shrimp, Eulimnadia texana. Journal of evolutionary biology. 27 (10), pp. 2080–2095.

WEEKS, S.C., BENVENUTO, C., SANDERSON, T.F., and DUFF, R.J., 2010. Sex chromosome evolution in the clam shrimp, Eulimnadia texana. Journal of evolutionary biology. 23 (5), pp. 1100–1106.

WEEKS, S.C., CHIARA BENVENUTO, and REED, S.K., 2006. When Males and Hermaphrodites Coexist: A Review of Androdioecy in Animals. Integrative and comparative biology. 46 (4), pp. 449–464.

WERREN, J.H., BALDO, L., and CLARK, M.E., 2008. Wolbachia: master manipulators of invertebrate biology. Nature reviews. Microbiology. 6 (10), pp. 741–751.

WILLING, E.-M., HOFFMANN, M., KLEIN, J.D., WEIGEL, D., and DREYER, C., 2011. Paired-end RAD-seq for de novo assembly and marker design without available reference. Bioinformatics . 27 (16), pp. 2187–2193.

WRIGHT, A.E., DEAN, R., ZIMMER, F., and MANK, J.E., 2016. How to make a sex

chromosome. Nature communications. 7, p. 12087.

ZAFFAGNINI, F. and TRENTINI, M., 1980. THE DISTRIBUTION AND REPRODUCTION OF TRIOPS CANCRIFORMIS (BOSC) IN EUROPE (CRUSTACEA NOTOSTRACA). Monitore Zoologico Italiano - Italian Journal of Zoology. 14 (1-2), pp. 1–8.

ZHOU, Q., ELLISON, C.E., KAISER, V.B., ALEKSEYENKO, A.A., GORCHAKOV, A.A., and BACHTROG, D., 2013. The Epigenome of Evolving Drosophila Neo-Sex Chromosomes: Dosage Compensation and Heterochromatin Formation. PLoS biology. 11 (11), p. e1001711.

ZHOU, Q., ZHANG, J., BACHTROG, D., AN, N., HUANG, Q., JARVIS, E.D., GILBERT, M.T.P., and ZHANG, G., 2014. Complex evolutionary trajectories of sex chromosomes across bird taxa. Science. 346 (6215), p. 1246338.

ZIEROLD, T., HANFLING, B., and GÓMEZ, A., 2007. Recent evolution of alternative reproductive modes in the 'living fossil' Triops cancriformis. BMC evolutionary biology. 7 (1), p. 161.

ZIEROLD, T., MONTERO-PAU, J., HÄNFLING, B., and GÓMEZ, A., 2009. Sex ratio, reproductive mode and genetic diversity in Triops cancriformis. Freshwater biology. 54 (7), pp. 1392–1405.

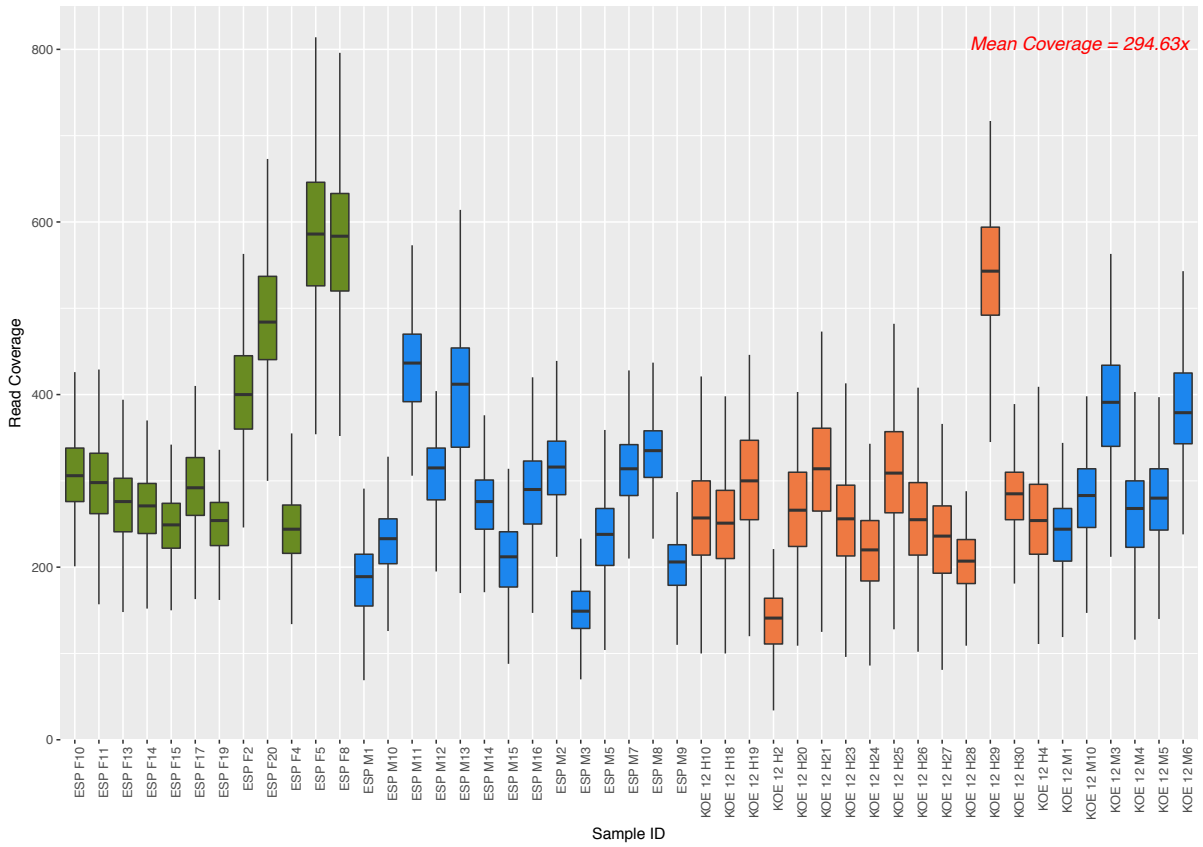
Appendix

Table 1. The individuals included in the RAD seq data with the location of population and sex.

Sample ID	RAD Library ID	Population	Sex
ESP_F2	Lib_2	Espolla, Spain	Female
ESP_F4	Lib_2	Espolla, Spain	Female
ESP_F5	Lib_2	Espolla, Spain	Female
ESP_F8	Lib_2	Espolla, Spain	Female
ESP_F10	Lib_2	Espolla, Spain	Female
ESP_F11	Lib_2	Espolla, Spain	Female
ESP_F13	Lib_2	Espolla, Spain	Female
ESP_F14	Lib_2	Espolla, Spain	Female
ESP_F15	Lib_3	Espolla, Spain	Female
ESP_F17	Lib_3	Espolla, Spain	Female
ESP_F19	Lib_3	Espolla, Spain	Female
ESP_F20	Lib_2	Espolla, Spain	Female
ESP_M1	Lib_4	Espolla, Spain	Male
ESP_M2	Lib_3	Espolla, Spain	Male
ESP_M3	Lib_4	Espolla, Spain	Male
ESP_M5	Lib_4	Espolla, Spain	Male
ESP_M7	Lib_3	Espolla, Spain	Male
ESP_M8	Lib_3	Espolla, Spain	Male
ESP_M9	Lib_3	Espolla, Spain	Male
ESP_M10	Lib_3	Espolla, Spain	Male
ESP_M11	Lib_3	Espolla, Spain	Male
ESP_M12	Lib_3	Espolla, Spain	Male
ESP_M13	Lib_3	Espolla, Spain	Male
ESP_M14	Lib_3	Espolla, Spain	Male
ESP_M15	Lib_4	Espolla, Spain	Male
ESP_M16	Lib_4	Espolla, Spain	Male
KOE_12_H2	Lib_1	Pond 12, Königswartha, Germany	Hermaphrodite
KOE_12_H4	Lib_1	Pond 12, Königswartha, Germany	Hermaphrodite
KOE_12_H10	Lib_1	Pond 12, Königswartha, Germany	Hermaphrodite

KOE_12_H18	Lib_1	Pond 12, Königswartha, Germany	Hermaphrodite
KOE_12_H19	Lib_1	Pond 12, Königswartha, Germany	Hermaphrodite
KOE_12_H20	Lib_1	Pond 12, Königswartha, Germany	Hermaphrodite
KOE_12_H21	Lib_1	Pond 12, Königswartha, Germany	Hermaphrodite
KOE_12_H23	Lib_1	Pond 12, Königswartha, Germany	Hermaphrodite
KOE_12_H24	Lib_1	Pond 12, Königswartha, Germany	Hermaphrodite
KOE_12_H25	Lib_1	Pond 12, Königswartha, Germany	Hermaphrodite
KOE_12_H26	Lib_1	Pond 12, Königswartha, Germany	Hermaphrodite
KOE_12_H27	Lib_1	Pond 12, Königswartha, Germany	Hermaphrodite
KOE_12_H28	Lib_2	Pond 12, Königswartha, Germany	Hermaphrodite
KOE_12_H29	Lib_2	Pond 12, Königswartha, Germany	Hermaphrodite
KOE_12_H30	Lib_2	Pond 12, Königswartha, Germany	Hermaphrodite
KOE_12_M1	Lib_4	Pond 12, Königswartha, Germany	Male
KOE_12_M3	Lib_4	Pond 12, Königswartha, Germany	Male
KOE_12_M4	Lib_4	Pond 12, Königswartha, Germany	Male
KOE_12_M5	Lib_4	Pond 12, Königswartha, Germany	Male
KOE_12_M6	Lib_4	Pond 12, Königswartha, Germany	Male
KOE_12_M10	Lib_4	Pond 12, Königswartha, Germany	Male

a)



b)

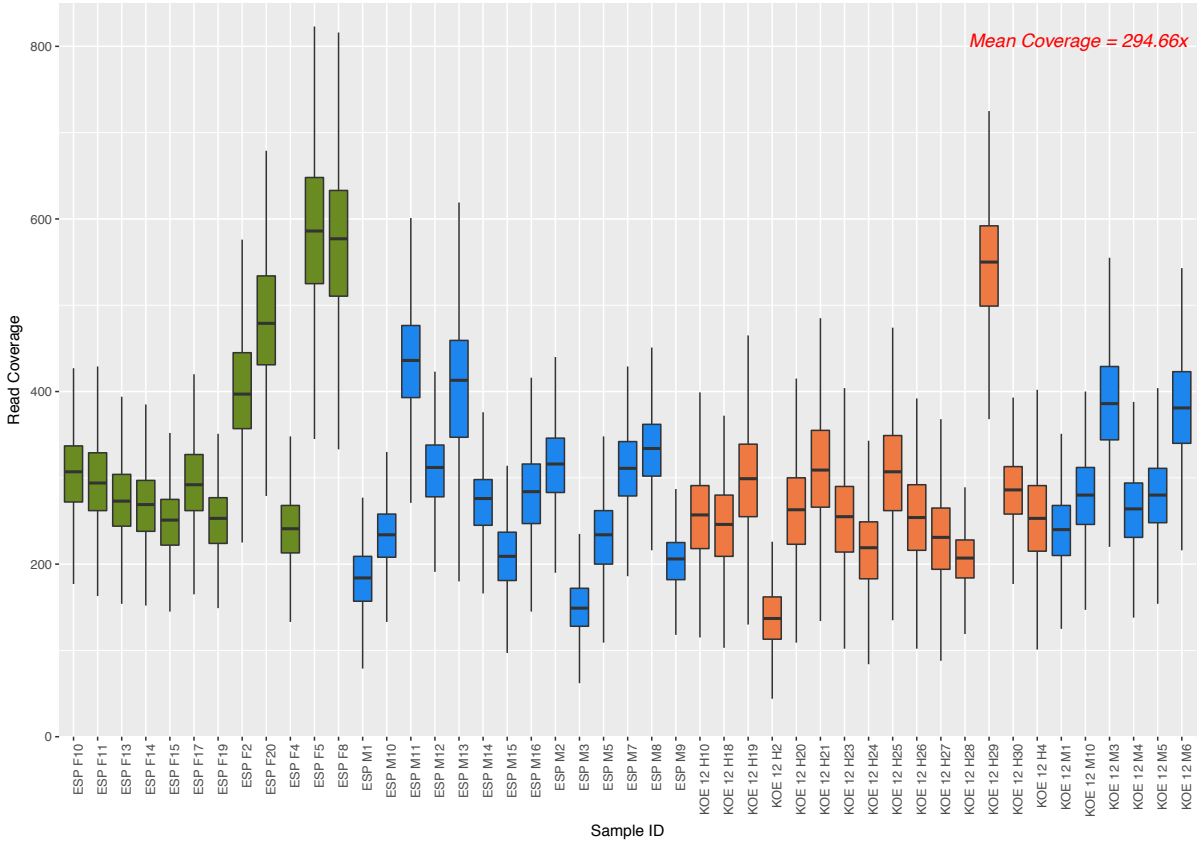


Figure 1. Average read coverage depths across individuals when mapping RAD sequence data to: a) male assembly, b) hermaphrodite assembly

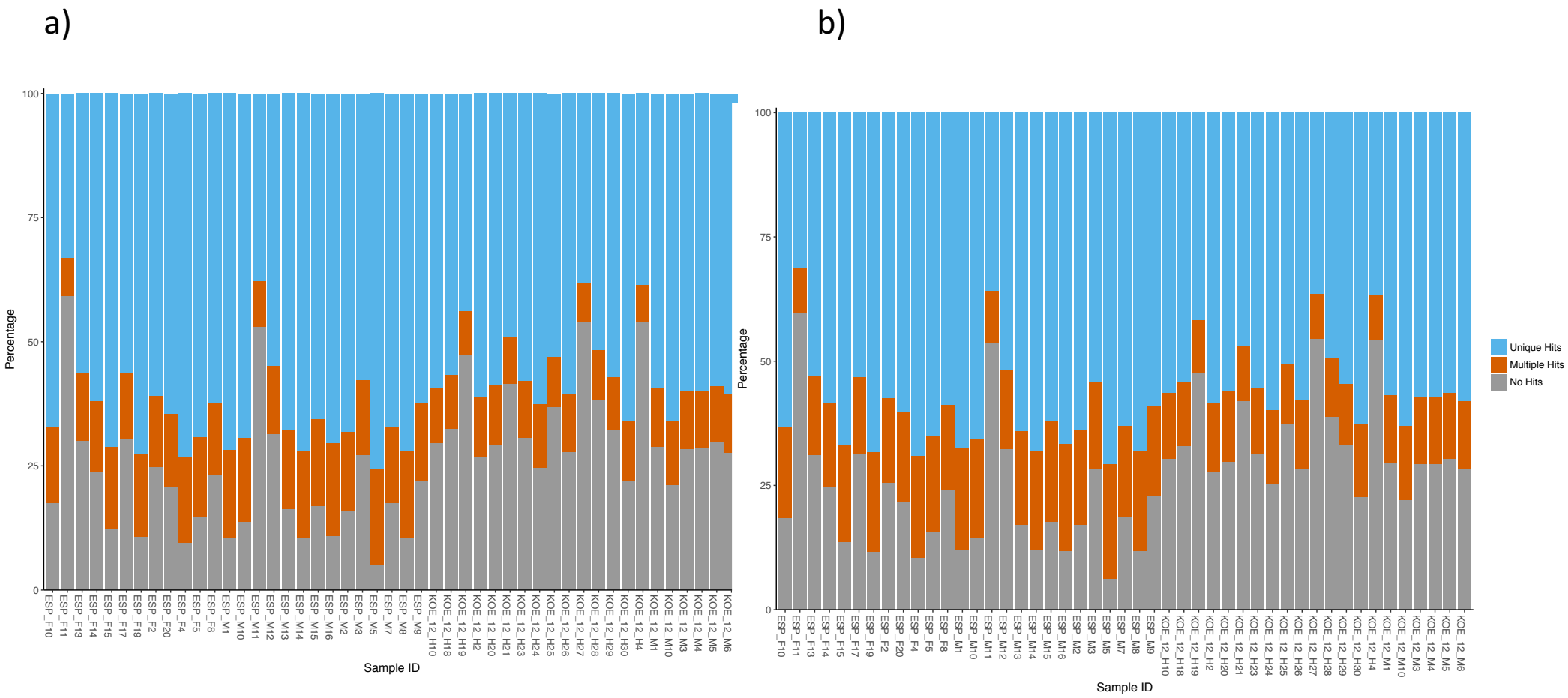


Figure 2. Mapping statistics when mapping RAD sequence data to: a) hermaphrodite assembly, b) male assembly. Unique hit is defined as one match on the assembly; multiple hit is defined as more than one match on the assembly and no matches.

Supplementary data available at DOI: 10.5281/zenodo.998272.