Heterozygosity at a conserved candidate sex determination locus is associated with female development in the clonal raider ant (*Ooceraea biroi*)

Kip D. Lacy[1*], Jina Lee[1], Kathryn Rozen-Gagnon[2,3], Wei Wang[3], Thomas S. Carroll[3], Daniel J.C. Kronauer[1,4*]

[1]Laboratory of Social Evolution and Behavior, The Rockefeller University, New York, NY, USA.

[2]Department of Molecular Genetics, University of Toronto, Toronto, Canada.

[3]Bioinformatics Resource Center, The Rockefeller University, New York, NY, USA.

[4]Howard Hughes Medical Institute, New York, NY, USA.

[*]Email: klacy@rockefeller.edu (K.D.L.); dkronauer@rockefeller.edu (D.J.C.K)

Abstract

Sex determination is a developmental switch that triggers sex-specific developmental programs. This switch is "flipped" by the expression of genes that promote male- or female-specific development. Many lineages have evolved sex chromosomes that act as primary signals for sex determination. However, haplodiploidy (males are haploid and females are diploid), which occurs in ca. 12% of animal species, is incompatible with sex chromosomes. Haplodiploid taxa must, therefore, rely on other strategies for sex determination. One mechanism, "complementary sex determination" (CSD), uses heterozygosity as a proxy for diploidy. In CSD, heterozygosity at a sex determination locus triggers female development, while hemizygosity or homozygosity permits male development. CSD loci have been mapped in honeybees and two ant species, but we know little about their evolutionary history. Here, we investigate sex determination in the clonal raider ant, *Ooceraea biroi*. We identified a 46kb candidate CSD locus at which all females are heterozygous, but most diploid males are homozygous for either allele. As expected for CSD loci, the candidate locus has more alleles than most other loci, resulting in a peak of nucleotide diversity. This peak negligibly affects the amino acid sequences of protein-coding genes, suggesting that heterozygosity of a non-coding genomic sequence triggers female development. This locus is distinct from the CSD locus in honeybees but homologous to a CSD locus mapped in two distantly related ant species, implying that this molecular mechanism has been conserved since a common ancestor that lived approximately 112 million years ago.

Introduction

Whether an animal develops as a male or a female is typically determined by a switch early in development. This switch, known as sex determination, is a multistep process beginning with primary sex determination signals triggering sex-specific splicing of downstream transcription factors that encode sexual identity (Williams and Carroll 2009). Although the downstream transcription factors are evolutionarily conserved, primary sex determination signals are evolutionarily labile and comprise a diverse range of environmental and genetic mechanisms (Bachtrog et al. 2014). Genetically encoded primary signals (sex-determining genes or chromosomes) are usually found in only one sex. Such systems do not work in haplodiploid taxa (an estimated 12% of animal species, including ants (Normark 2003)), where any allele on any homologous chromosome can be transmitted from diploid females to their haploid male sons, and all alleles experience selection to be viably transmitted from both sexes (Whiting 1935). Haplodiploids, therefore, require alternative sex determination mechanisms, but we know little about the molecular details.

Some parasitoid wasps use a mechanism involving maternal imprinting (Verhulst et al. 2010; Zou et al. 2020). However, in many Hymenoptera, inbred crosses produce diploid males, which led to the hypothesis that female development is triggered by heterozygosity as a proxy for diploidy (Whiting 1933; Whiting 1943). Under this “complementary sex determination” (CSD) hypothesis, different alleles at a given sex determination locus “complement” one another, meaning that heterozygosity triggers female development (**Fig. 1A**). By contrast, the presence of a single allele (either because the individual is haploid or because a diploid individual is homozygous at the sex determination locus) permits male development. Under CSD, diploid males produce diploid sperm and, therefore, are functionally sterile (van

Wilgenburg et al. 2006). This has several evolutionary consequences (Cook and Crozier 1995); at the population level, there is selection against homozygosity, often meaning obligatory avoidance of inbreeding (Page 1980). At the gene level, there is negative frequency-dependent (balancing) selection on CSD loci, resulting in the co-occurrence of many alleles at similar frequencies within populations (Laidlaw et al. 1956).

Having been a favored hypothesis for more than half a century, CSD was demonstrated in the honeybee *Apis mellifera* in 2003 (Beye et al. 2003). In *A. mellifera*, biallelic heteromers of the protein encoded by the *complementary sex determiner* (*csd*) gene trigger female development, whereas monoallelic homomers permit male development (Beye et al. 2003; Beye et al. 2013; Otte et al. 2023; Seiler and Beye 2024). Although this primary sex determination signal differs from primary signals in other insects, it nonetheless converges upon the pathway that transduces sex determination signals in holometabolous insects (Bopp et al. 2014; Sawanth et al. 2016; Wexler et al. 2019). The Csd protein sex-specifically splices the closely linked splicing factor *feminizer* (*fem*) (the honeybee version of *transformer* (*tra*) in *Drosophila* and other insects), which in turn leads to the production of sex-specific isoforms of *doublesex* (*dsx*), a transcription factor that regulates sex-specific development (Hasselmann et al. 2008; Gempe et al. 2009; Verhulst et al. 2010; Zou et al. 2020). Intriguingly, *csd* is a paralog of *fem*. Following the prediction that CSD loci should evolve under balancing selection, many *csd* alleles are held at similar frequencies, and the locus bears signatures of adaptive evolution (Hasselmann and Beye 2004), with amino acid-level heterozygosity in certain protein domains required for female development (Otte et al. 2023).

CSD is assumed to occur in many Hymenoptera that occasionally produce diploid males (van Wilgenburg et al. 2006). Until recently, however, heterozygosity-dependent female

development had not been demonstrated outside of *A. mellifera*, which remained the only species for which a CSD locus had been genetically mapped. A recent genetic mapping study in the wasp *Lysiphlebus fabarum* identified at least one and as many as four candidate sex determination loci (Matthey-Doret et al. 2019) (the multi-locus extension of the CSD hypothesis proposes that heterozygosity at any of multiple sex determination loci is sufficient to trigger female development (Crozier 1971) (**Fig. 1B**)). However, a requirement of heterozygosity at these loci for female development was not demonstrated. In the ant *Vollenhovia emeryi*, a mapping study found two sex determination QTL (Miyakawa and Mikheyev 2015), with follow-up work suggesting that these function as a multi-locus CSD system (Miyakawa and Miyakawa 2023). One of these QTL (*V.emeryiCsdQTL1*) contains two closely linked *tra* homologs, reminiscent of the locus containing *csd* and *fem* in *A. mellifera*. By contrast, the other QTL (*V.emeryiCsdQTL2*) only contains genes without annotated functions in sexual development. Although *V. emeryi* sex determination is likely integrated through *tra* and *dsx*, the molecular basis of their primary sex determination signals remains unclear (Miyakawa et al. 2018; Miyakawa and Miyakawa 2023).

Recently, a sex determination locus was identified in the Argentine ant, *Linepithema humile* (Pan et al. 2024). Consistent with CSD, all pairwise heterozygous combinations of the seven identified alleles were found in females, whereas diploid males were invariably homozygous for one of the seven alleles. Moreover, these alleles occurred at roughly equal frequencies within populations, suggesting evolution under balancing selection. The mapped locus is located within a noncoding genomic region tightly linked to *ant noncoding transformer splicing regulator* (*ANTSR*), a long noncoding RNA (lncRNA) with previously uncharacterized function. *ANTSR* knockdown in female-destined embryos leads to male-specific splicing of *tra*,

suggesting that this lncRNA transduces the primary sex determination signal encoded by the heterozygosity (or hemi- or homozygosity) of the linked noncoding region (Pan et al. 2024). Because the synteny of this region (including the presence of a lncRNA putatively homologous to *ANTSR*) is conserved across the ants, bees, and vespoid wasps, it was suggested that this sex determination locus may be deeply conserved among the Aculeata (Pan et al. 2024). However, synteny is insufficient to demonstrate functional conservation, and genetic mapping of sex determination has yet to be performed in most of these taxa.

To explore the evolution of sex determination across ants, we investigated the topic in the clonal raider ant, *Ooceraea biroi*. Diploid males occur sporadically in this species, suggesting that, like other ants, it might employ CSD (Kronauer et al. 2012). *O. biroi* reproduces via a mode of parthenogenesis in which two haploid nuclei from a single meiosis fuse to produce diploid offspring (central fusion automixis) (Oxley et al. 2014). Although heterozygosity is maintained with high fidelity despite meiotic crossover recombination (Lacy et al. 2024), losses of heterozygosity occur occasionally, probably due to the inheritance of one recombined and one non-recombined version of a homologous chromosome (Kronauer et al. 2012; Oxley et al. 2014; Trible et al. 2023; Lacy et al. 2024). Therefore, if *O. biroi* uses CSD, diploid males might result from losses of heterozygosity at sex determination loci (**Fig. 1C**), similar to what is thought to occur in other asexual Hymenoptera that produce diploid males (Rabeling and Kronauer 2012; Matthey-Doret et al. 2019). Here, we use whole genome sequencing to map a sex determination locus, demonstrate that most diploid males carry a loss of heterozygosity at this locus, assess the homology of this locus with other sex determination loci, and explore whether balancing selection has shaped evolution at this locus.

Results

Identification of a candidate sex determination locus on chromosome 4

To map sex determination loci in *O. biroi*, we first identified males based on sexual dimorphism and distinguished diploid from haploid males based on heterozygosity at eight genetic markers (**Supplementary Table S1, File S1, and Fig. S1**). We then sequenced the genomes of 16 diploid males and compared them with 19 previously sequenced genomes of diploid females (Trible et al. 2023; Lacy et al. 2024) (descriptions and metadata for all genomes sequenced in this study are in **Supplementary Table S2**). CSD loci must be heterozygous to trigger female development but can be homozygous in diploid males. To identify such sites, we calculated a "CSD index" for each single nucleotide polymorphism (SNP), which equals zero if any female is homozygous and equals the proportion of diploid males that are homozygous if all females are heterozygous. We found a peak of the CSD index on chromosome 4, which deviated significantly from the null distribution of CSD index values (**Fig. 2A**). At this peak, homozygosity levels differed significantly between males and females (**Supplementary Fig. S2**).

To corroborate our results, we looked for evidence of segmental losses of heterozygosity along chromosome 4. Eleven out of 16 sequenced diploid males bore segmental losses of heterozygosity on chromosome 4 (**Fig. 2B**). The intersection of these LOH segments comprised a 46kb peak at which all females were heterozygous, and most (11 out of 16) diploid males were homozygous. Under the CSD hypothesis, homozygosity for any allele at CSD loci permits male development. Consistently, *O. biroi* diploid males were homozygous for either allele at the 46kb region on chromosome 4 (**Table 1, Supplementary Fig. S3**).

The stretches of homozygosity found in diploid males could have resulted from segmental deletions resulting from improperly repaired DNA damage or rare copy-neutral losses of heterozygosity resulting from thelytokous parthenogenesis (Kronauer et al. 2012; Oxley et al. 2014; Trible et al. 2023; Lacy et al. 2024). To distinguish between these two scenarios, we analyzed read depth and found that runs of homozygosity were not accompanied by changes in copy number (**Supplementary Fig. S3**). This implies that diploid males arise when rare losses of heterozygosity, which result from crossover recombination during meiosis, span the CSD locus.

Surprisingly, five of the 16 diploid males were not homozygous in this 46kb region. One possible explanation is that these individuals had mutations in other genes involved in the sex-determination pathway. We therefore inspected all unique mutations and losses of heterozygosity in these five diploid males, but did not find any in genes with annotated sex determining functions (**Supplementary File S2**). These individuals may carry mutations or losses of heterozygosity in genes or regulatory elements with unannotated sex determination function. Alternatively, male development may have been triggered by rare stochastic perturbations to gene expression or splicing. We address these possibilities further in the Discussion.

<u>*O. biroi*'s candidate CSD locus is homologous to another ant sex determination locus but not to honeybee *csd*</u>

To assess homology to CSD loci mapped in other species, we identified *O. biroi* orthologs of the genes found in other mapped sex determination loci. We found no homology between the genes within the *O. biroi* CSD index peak and any of the genes within the putative *L. fabarum* CSD loci (**Supplementary Table S3**). By identifying orthologs of genes within the two sex determination QTL identified in *V. emeryi* (Miyakawa and Mikheyev 2015)

(**Supplementary Table S4**), we also identified homology to the *L. humile* CSD locus, because a subset of *V.emeryiCsdQTL2* is homologous to the CSD identified in *L. humile* (Pan et al. 2024). Homology to *V.emeryiCsdQTL2* was scattered across one arm of *O. biroi* chromosome 4, including the 46kb CSD index peak, and homology to the *L. humile* CSD locus mapped within the *O. biroi* CSD index peak (**Fig. 3A**). The *O. biroi* CSD index peak contains four protein-coding genes that are also present in *V.emeryiCsdQTL2*, including *HCF*, *COPA*, an uncharacterized gene, and *CRELD2*. Notably, the candidate CSD region in *O. biroi* also includes a large stretch of sequence with no annotated genes downstream of an uncharacterized lncRNA. LncRNAs often retain their functions and synteny with surrounding genes despite divergence in sequence homology (Chodroff et al. 2010; Ulitsky et al. 2011; Quinn et al. 2016), raising the possibility that, although this lncRNA has limited sequence homology with *L. humile ANTSR*, it may have a homologous function (Pan et al. 2024). Regardless of the underlying molecular mechanism, our data suggest that this CSD locus is conserved between *O. biroi*, *L. humile*, and *V. emeryi* (**Fig. 3B**). This would mean that this locus is ancient, as these three species belong to different ant subfamilies that diverged roughly 112 million years ago (Borowiec et al. 2025).

We performed our genetic mapping in clonal line A because this was the only clonal line in which we found diploid males (see below). However, different clonal lines of *O. biroi* vary in which portions of the genome are ancestrally homozygous (Oxley et al. 2014). Therefore, it was unclear whether heterozygosity at the mapped CSD locus is required for female development in all clonal lines of *O. biroi*. To investigate this, we inspected heterozygosity in diploid females from six different clonal lines by retrieving previously published genome sequences from clonal line B (Lacy et al. 2024) and sequencing genomes of diploid females from four additional clonal lines (from both the native and invasive ranges of *O. biroi* (Kronauer et al. 2012; Trible et al.

2020)). All females from all six clonal lines (including 26 diploid females from clonal line B) were heterozygous at the CSD index peak, consistent with its putative role as a CSD locus in all *O. biroi* (**Fig. 3C**).

The CSD index peak has high diversity in a non-coding region.

Because CSD loci are expected to evolve under balancing selection, we expect many different alleles to exist at those loci. For example, as many as 19 *csd* alleles were found to segregate in honeybee populations (Hasselmann et al. 2008), and seven alleles were found to segregate in a population of *L. humile*, despite a recent genetic bottleneck (Pan et al. 2024). Looking for elevated genetic diversity is impossible by only studying genomes from a single clonal line (i.e., asexual descendants of a single diploid female), so we looked across diploid females from six different clonal lines. We calculated nucleotide diversity (π) in 5kb sliding windows (step size = 1kb) across the genome and found a nucleotide diversity peak that fell within the CSD index peak (**Fig. 4A, Supplementary Fig. S4**), between positions 1,717,000 and 1,730,000 on chromosome 4 (**Fig. 4B**).

We hypothesized that this nucleotide diversity peak resulted from many alleles found among clonal lines of *O. biroi* at this locus. Identifying these alleles requires DNA sequences of individual haplotypes, whereas short-read genome sequences of diploid females only provide diploid genotypes. To investigate allelic diversity, we retrieved previously published haploid male genome sequences from clonal lines A and B (Lacy et al. 2024) and sequenced whole genomes of haploid males from additional clonal lines. We assembled their genomes *de novo* to identify the different haplotypes found in this region. In addition to the haplotype found in the reference genome, we identified six additional haplotypes (or “alleles”) in the five clonal lines

for which we sequenced at least one haploid male genome (**Fig. 4C**, **Table 2**). These alleles differ substantially in their DNA sequences in the region with high nucleotide diversity (**Fig. 4C**).

To investigate which genetic elements might act to determine sex in a heterozygosity-dependent manner, we inspected all annotated genes in the CSD index peak (**Fig. 4D**). Within this region, none of the protein-coding genes contained substantial heterozygosity that affected amino acid identity (**Supplementary File S3**). To determine whether any genes or exons were missing from our annotation, we performed RNA sequencing of embryos and several other life stages (including female and male adults) using several different technologies (see Methods). Briefly, we used long-read RNA sequencing (PacBio IsoSeq) to extend gene models and used several different library preparation methods followed by standard short-read (Illumina) sequencing: standard poly-A tail library prep to capture mRNAs, ribosomal RNA depletion to capture lncRNAs that are not poly-adenylated, and small RNA sequencing to annotate miRNAs, endo-siRNAs, and putative piRNAs. Although this improved the *O. biroi* genome annotation by adding new genes and improving existing gene models (**Supplementary Table S5, File S4**), no new genes or exons were identified near the CSD index peak.

To look for sex-biased gene expression and splicing, we performed differential gene expression and differential transcript usage analysis between the transcriptomes of three male and three female early-stage (white) pupae. We used pupae for this experiment because this is the earliest life stage at which males and females can be readily distinguished. Primary sex determination signals act early during embryonic development (Sawanth et al. 2016) and thus need not be differentially expressed later in development. However, these genes can still be expressed as late as the pupal stage (Beye et al. 2003). Many genes were differentially expressed

between males and females (24.6% of expressed genes) or had one or more differentially used exons, including four genes near the CSD index peak (LOC105285605, LOC105283850, LOC105283849, and LOC105283844) (**Supplementary Tables S5 and S6; Fig. S5**). However, none of these overlapped with the region of elevated genetic diversity. Very few reads from the male and female pupae samples aligned to the lncRNA, and although more reads aligned in RNA-seq libraries from eggs prepared using ribosomal RNA depletion, these data are insufficient for a rigorous assessment of sex-specific expression at this locus (**Supplementary Fig. S6**). Therefore, it remains unclear whether this lncRNA plays a role in sex determination in *O. biroi*.

The peak of nucleotide diversity minimally affects amino acid sequences and is primarily located in a non-coding genomic region (**Fig. 4B-D**). Thus, a non-coding genetic element within this region may instruct sex determination in *O. biroi* in a heterozygosity-dependent manner. This is similar to *L. humile*, where the mapped CSD locus had high variability in non-coding and non-exonic regions (Pan et al. 2024). RNA interference of the nearby lncRNA, *ANTSR*, led to male-specific splicing of *tra*, raising the possibility that heterozygosity or homozygosity in the non-exonic and non-coding *L. humile* CSD locus affects the expression level of *ANTSR*, which instructs sex-specific splicing of *tra* (Pan et al. 2024). Because the peak of nucleotide diversity within our CSD index peak minimally affects protein-coding sequences and the region is closely linked to a lncRNA that is putatively an ortholog of *L. humile ANTSR*, our data suggest that the mechanism of CSD in *O. biroi* may be conserved with *L. humile*.

<u>Diploid male production differs across *O. biroi* clonal lines.</u>

Over the course of our study, we sampled many males from clonal lines A and B, and a few males from four additional clonal lines. Despite the fact that we maintain more colonies of clonal line B than of clonal line A in the lab, all the diploid males we detected came from clonal line A. We, therefore, looked back at data from previously genotyped males and found that this pattern held across a previous study (Kronauer et al. 2012). In total, from clonal line A, 22 out of 114 males were diploid, whereas all 78 males genotyped from clonal line B were haploid (**Table 3**). This disparity is statistically significant (χ^2=15.4, df=1, p<0.0001), suggesting that the two clonal lines differ in their propensity to produce diploid males.

One possible explanation for this phenomenon could be that *O. biroi* has multi-locus CSD (**Fig. 1B**), but that clonal line A is ancestrally homozygous for all but one of the loci, whereas clonal line B retains heterozygosity at two or more loci. Because multi-locus CSD requires homozygosity at all sex determination loci for diploids to develop as males, in this scenario, two or more losses of heterozygosity would be needed for diploid males to develop in clonal line B. In contrast, only one loss of heterozygosity would be required in clonal line A, which we used for our genetic mapping. However, we identified no additional candidate loci in our mapping study, and other possibilities are equally likely, including that clonal lines A is more prone to loss of heterozygosity, or is more permissive to diploid male development than other clonal lines.

<u>*O. biroi* lacks a *tra*-containing CSD locus.</u>

A second QTL region identified in *V. emeryi* (*V.emeryiCsdQTL1*) contains two closely linked *tra* homologs, similar to the closely linked honeybee *tra* homologs, *csd* and *fem* (Miyakawa and Mikheyev 2015). This, along with the discovery of duplicated *tra* homologs that

undergo concerted evolution in bumblebees and ants (Schmieder et al. 2012; Privman et al. 2013) has led to the hypothesis that the function of *tra* homologs as CSD loci is conserved with the *csd*-containing region of honeybees (Schmieder et al. 2012; Miyakawa and Mikheyev 2015). However, other work has suggested that *tra* duplications occurred independently in honeybees, bumblebees, and ants (Hasselmann et al. 2008; Koch et al. 2014), and it remains to be demonstrated that either of these *tra* homologs acts as a primary CSD signal in *V. emeryi*. To test whether the *tra* homolog-containing QTL (*V.emeryiCsdQTL1*) might be involved in CSD in *O. biroi*, we searched for homology to this QTL in the *O. biroi* genome and identified several regions scattered across *O. biroi* chromosome 2 (**Fig. 5A**). However, we did not find a CSD index peak on chromosome 2, and no sex determination locus with homology to the *V. emeryi* QTL was found in *L. humile* (Pan et al. 2024). Thus, the function of *V.emeryiCsdQTL1* as a primary sex determination signal might have evolved after this lineage separated from other ants (**Fig. 5B**). Alternatively, the sex determination activity of *V.emeryiCsdQTL1* could be ancestral to these three ant species and independently lost in *O. biroi* (at least in clonal line A) and *L. humile*.

Intriguingly, while clonal lines B, D, I, and M retain heterozygosity at all sites with homology to *V.emeryiQTL1* (**Fig. 5C**), clonal lines A and C are homozygous in a region homologous to a portion of *V.emeryiQTL1*. We note, however, that heterozygosity in this region is limited, without elevated nucleotide diversity (**Supplementary Fig. S4**), and there are only two nonsynonymous heterozygous variants in this region, with clonal line B conspicuously lacking any nonsynonymous heterozygosity. Therefore, it seems unlikely that this region bears a CSD locus in *O. biroi*. We also note that all clonal lines are heterozygous in the region containing *O. biroi*'s *tra* ortholog (**Fig. 5C**), and that this gene is sex-specifically spliced

(**Supplementary Fig. S7**), similar to *tra* homologs in other Holometabola (Hasselmann et al. 2008; Miyakawa and Miyakawa 2023). This contrasts with the *tra-like* gene in *V. emeryi*, which, like *A. mellifera csd*, is not sex-specifically spliced (Miyakawa and Miyakawa 2023).

Discussion

Genetically mapping sex determination loci in diverse taxa is key to understanding how sex determination evolves. Here, we investigated the mode of sex determination in the clonal raider ant, *O. biroi*. First, we mapped a candidate sex determination locus that was heterozygous in all females, but homozygous for either allele in most diploid males (**Fig. 2**). Next, we showed that most diploid males bear rare losses of heterozygosity that arise during thelytokous parthenogenesis (**Fig. 2**). We then showed that this locus is homologous to a sex determination locus found in two other ant species—*V. emeryi* and *L. humile* (**Fig. 3**), suggesting that this mechanism of sex determination may be conserved among formicoid ants. By sequencing whole genomes of haploid males and diploid females from other clonal lines of *O. biroi*, we showed that the mapped locus corresponds to a region with high genetic diversity, following the expectation that CSD loci should evolve under balancing selection (**Fig. 4**). This diversity peaks in a noncoding genomic region, and we found limited evidence for functional heterozygosity in nearby protein-coding sequences, suggesting that heterozygosity in this noncoding sequence may trigger female development. Finally, we demonstrated that *O. biroi* CSD in clonal line A does not map to the *tra*-containing sex determination locus previously identified in *V. emeryi*, but it remains unclear whether *O. biroi* ancestrally has single-locus or multi-locus CSD (**Fig. 5**).

The presence of diploid males in many ant taxa previously led to the conclusion that most ants employ CSD (van Wilgenburg et al. 2006). Along with *L. humile* (Pan et al. 2024) and *V.*

emeryi (Miyakawa and Mikheyev 2015), *O. biroi* is now one of three ant species for which heterozygosity-dependent female development has been demonstrated at mapped CSD loci. Our mapped locus appears to be homologous to *V.emeryiCsdQTL2* and to the locus identified in *L. humile* (**Fig. 3**). As was found in *L. humile*, this locus colocalizes with a peak of nucleotide diversity (**Fig. 4**), with seven alleles found among five clonal lines (**Fig. 4C, Table 2**), consistent with the expectation that CSD loci should harbor many alleles due to evolution under balancing selection. This implies that the role of this locus in CSD evolved before the divergence of the Dorylinae from the other formicoid ants roughly 112 million years ago (Borowiec et al. 2025). Although synteny analyses raised the possibility that this locus arose early within the evolution of the Aculeata (Pan et al. 2024), genetic mapping studies of additional representative species of different ant subfamilies (especially in the poneroid clade and the Leptanillinae) and across other Aculeata will be required to determine when the function of this locus in CSD originated and how it is distributed across taxa.

In *L. humile*, a lncRNA (*ANTSR*) closely linked to this locus plays a role in sex determination (Pan et al. 2024). Although we observed some expression of the putatively homologous lncRNA in *O. biroi* eggs of unknown sex, we were unable to assess sex-specific expression. The reason is that males in *O. biroi* are exceedingly rare and morphologically only identifiable at the pupal stage, well after primary sex determination signals are transduced to self-sustaining sex-specific splicing of downstream transcription factors during the early stages of embryonic development (Gempe and Beye 2011). Therefore, we likely missed differential gene expression relevant to sex determination that occurs before the pupal stage.

We did not find a peak of the CSD index at the *O. biroi tra* homolog (**Fig. 5**), indicating that, as in *L. humile*, the mode of CSD is not homologous to the system in *A. mellifera* mediated

by the *transformer* homologs *csd* and *fem*. This was previously suspected for *O. biroi* due to the presence of only a single *tra* homolog (Oxley et al. 2014) rather than the two closely linked *tra* homologs found in other ants (Privman et al. 2013). In *V. emeryi*, however, one of the two Csd QTLs (*V.emeryiCsdQTL1*) spans the *tra* locus, raising the possibility that the molecular mechanism of CSD might, in part, be conserved between *A. mellifera* and *V. emeryi* (Miyakawa and Mikheyev 2015). It has not been demonstrated that either of *V. emeryi*'s *tra* homologs act as a primary sex determination signal. Our results raise the possibility that the role of *V.emeryiCsdQTL1* as a primary sex determination signal evolved independently of *A. mellifera csd* and after *V. emeryi* diverged from *L. humile*. However, further investigations into the mechanisms of sex determination in *V. emeryi* and other ants are required to reconstruct the evolutionary history of *V.emeryiCsdQTL1'*s function in hymenopteran sex determination.

Curiously, diploid males were only produced in clonal line A, even though many males were collected from other clonal lines (**Table 3**). One possible explanation is that *O. biroi* has multi-locus CSD, and clonal line A is ancestrally homozygous for all but one of the loci. Multi-locus CSD has been suggested to limit the extent of diploid male production in asexual species under some circumstances (Vorburger 2013; Matthey-Doret et al. 2019). Clonal line A is ancestrally homozygous for a portion of the region homologous to the second *V. emeryi* QTL, although it retains heterozygosity in the region overlapping *tra* (**Fig. 5C**). However, this region lacks elevated genetic diversity and functional heterozygosity, making it a poor candidate for a second, undetected, CSD locus. Further study would be required to determine whether *O. biroi* has a second CSD locus. However, there may be other explanations for diploid males being found exclusively in clonal line A. For example, clonal line A could be more likely to lose

heterozygosity than other clonal lines, or diploid males in clonal line A could be more likely to develop and survive. Currently, we have no data to distinguish between these hypotheses.

Another peculiarity of our study is that five of the 16 sequenced diploid males retained heterozygosity at the mapped CSD locus. One possible explanation for this would be that our mapped locus is, in fact, a false positive. However, this seems unlikely given that 1) those losses of heterozygosity occurred independently in multiple genetic backgrounds within clonal line A (**Supplementary Fig. S3**), 2) a peak of nucleotide diversity co-occurs with the mapped locus (**Fig. 4**), and 3) this locus is homologous to CSD loci mapped in two other ant species (**Fig. 3**). Stochastic gene expression or splicing fluctuations may have caused these individuals to develop as males despite having female genotypes. Certain allelic combinations of *A. mellifera csd* trigger female development with incomplete penetrance (Beye et al. 2013), and the allelic combination found in clonal line A might similarly be permissive to occasional male development. Alternatively, these diploid males could result from mutations in or losses of heterozygosity at different genetic loci. We note that it might be easier to find rare stochastically produced diploid males in laboratory colonies of *O. biroi* than in sexual species. This is because genetically encoded diploid males only result from loss of heterozygosity, which is rare in *O. biroi* (Oxley et al. 2014; Lacy et al. 2024). By contrast, in sexually reproducing populations, diploid males homozygous for sex determination loci will arise much more frequently due to matings between individuals that share an allele at the sex determination locus (this effect is exacerbated in invasive species like *L. humile* that have undergone recent genetic bottlenecks). Thus, if stochastically produced diploid males occur at very low baseline levels in ants, such males will rarely be detected in sexual species. In *O. biroi*, however, they could make up a

substantial proportion of diploid males, simply because diploid males caused by homozygosity at the CSD locus are likewise extremely rare.

Finally, we note that heterozygosity-dependent female development, a fundamental tenet of CSD, has major implications for the evolution of breeding systems and reproductive modes. Due to the high fitness cost of producing sterile diploid males, species with CSD should obligatorily outbreed or have other mechanisms of maintaining heterozygosity (van Wilgenburg et al. 2006). In this study, most diploid males resulted from rare crossover-associated losses of heterozygosity that spanned the putative CSD locus (the shortest loss of heterozygosity at the mapped locus was 46kb and, therefore, likely resulted from crossover recombination rather than gene conversion). Intriguingly, in many taxa, asexual lineages reproduce via mechanisms that lead to complete homozygosity in a single generation (Suomalainen et al. 1987; Ma and Schwander 2017). However, among the social Hymenoptera (and other Vespoids), the known modes of asexual reproduction all maintain at least some heterozygosity (Rabeling and Kronauer 2012; Ma and Schwander 2017). For formicoid ants, bees, and other taxa with CSD, asexual reproduction with high rates of heterozygosity loss would incur a steep fitness penalty due to the production of sterile diploid males. This may help explain the unusual inheritance system recently described in *O. biroi*, where following crossover recombination, reciprocally recombined chromatids are faithfully co-inherited so that heterozygosity is rarely lost (Lacy et al. 2024).

How this ancient putative CSD locus triggers female development in a heterozygosity-dependent manner remains unknown. In *L. humile*, RNAi experiments suggested that the expression level of the lncRNA *ANTSR* affects sex-specific splicing of *tra*, even though heterozygosity in the exons of *ANTSR* is not required for this effect (Pan et al. 2024). Thus, the

heterozygosity-dependent function of the candidate CSD locus may depend on local regulatory interactions between homologous chromosomes, similar to a phenomenon known as transvection (Duncan 2002). To better understand how heterozygosity induces female development in ants, future studies will need to fine-map and functionally characterize this causal locus.

Materials and Methods

Software Versions

The versions of all software used for analyses are provided in **Supplementary Table S8**.

Identifying Male Ploidy

Clonal raider ant colonies are composed of female ants that reproduce asexually via a type of thelytokous parthenogenesis termed central fusion automixis (Oxley et al. 2014). Males are produced only sporadically (Oxley et al. 2014). Females of this species are diploid, whereas males can either be haploid (presumably arising via a failure of central fusion following female meiosis) or diploid (presumably arising due to rare losses of heterozygosity at CSD loci). Males are easily recognizable due to substantial sexual dimorphism (males have eyes and wings, whereas females are blind and flightless, and males also have different body shape and more darkly pigmented cuticles than females). We sampled them opportunistically whenever they were observed in colonies. We distinguished diploid males from haploid males based on heterozygosity at several unlinked genetic markers (**Supplementary Table S1**). For this, we disrupted one leg from each male using a Qiagen TissueLyser II and extracted DNA using Qiagen's QIAmp DNA Micro Kit. We then PCR-amplified each marker and genotyped using Sanger sequencing (for most markers) or via gel electrophoresis either after restriction digestion

(for two markers) or directly after amplification (for one marker at which the different alleles are different sizes). Details of the genotyping methods can be found in **Supplementary Table S1**.

Genome Sequencing and Preliminary Analysis

For each library, we disrupted an individual ant with a metal bead and Qiagen's TissueLyser II and extracted DNA with Qiagen's QIAmp DNA Micro Kit. We prepared Nextera Flex Illumina DNA sequencing libraries for 16 diploid males from clonal line A, six diploid females from different clonal lines, and seven haploid males from different clonal lines. We sequenced these libraries using an Illumina NovaSeq 6000. We also accessed libraries for individual genomes from previous studies. The information for all DNA sequencing libraries used in this study can be found in **Supplementary Table S2**.

We trimmed reads using Trimmomatic 0.36 (Bolger et al. 2014), aligned them to the *O. biroi* reference genome (McKenzie and Kronauer 2018) (Obir_v5.4, GenBank assembly accession: GCA_003672135.1) using bwa mem (Li 2013), and then sorted, deduplicated, and indexed using picard (http://broadinstitute.github.io/picard/). We called variants using GATK HaplotypeCaller (version 4.2) (Poplin et al. 2018). To exclude falsely collapsed regions in the reference genome, we filtered against sites found to be heterozygous in haploid males. We also screened out any variants for which putatively heterozygous samples had a proportionate minor allelic depth less than 0.25 and/or for which putatively homozygous samples had a minor allelic depth greater than zero. We performed custom filtering of variants and calculated statistics using pyvcf 0.6.8 (https://github.com/jamescasbon/PyVCF), and performed sliding window analyses using pybedtools 0.9.0 (Dale et al. 2011).

CSD Index Mapping

CSD loci are heterozygous in females but can be homozygous in diploid males. Thus, traditional association mapping cannot identify CSD loci because particular alleles do not trigger female development. Therefore, to map candidate CSD loci, we used a "CSD Index," which equals zero if any female is homozygous, but if all females are heterozygous, it equals the proportion of diploid males that are homozygous. Before calculating this index, we had to identify ancestrally heterozygous sites. All individuals within an *O. biroi* clonal line are descended asexually from a single diploid female ancestor, and CSD loci would have been heterozygous in that common ancestor. Putatively ancestrally heterozygous variants are those for which at least one diploid individual is heterozygous, and all other diploid individuals are either heterozygous for the same two alleles, or homozygous for one of the two alleles for which other diploid(s) are heterozygous (Oxley et al. 2014; Lacy et al. 2024). After identifying putatively ancestrally heterozygous SNPs, we calculated the CSD index for each SNP.

To assess the likelihood that our observed CSD index values were due to chance (false positives), we randomly shuffled the identities of males and females to generate a null distribution of CSD index values for each SNP. In doing so, we maintained the underlying genetic structure by only shuffling sex labels within colonies. We computed p-values by calculating the proportion of permutations with values greater than or equal to the observed CSD index. To correct for multiple testing and control the false discovery rate across the genome-wide analysis, we applied the Benjamini-Hochberg False Discovery Rate correction and then recorded the mean CSD index value in sliding windows.

To directly evaluate the statistical association between homozygosity and sex, we applied Fisher's Exact Test to assess whether SNP homozygosity frequencies differed significantly

between males and females and used the Benjamini-Hochberg False Discovery Rate correction to adjust p-values.

Losses of Heterozygosity

We identified losses of heterozygosity as ancestrally heterozygous SNPs that had become homozygous in a given sample. To assess whether homozygosity for either allele at the mapped region permitted male development, we randomly chose a reference haploid male and assigned allelic identity based on whether a focal sample was homozygous for the same allele or the alternate allele to that possessed by the reference haploid male. To clearly illustrate losses of heterozygosity, we used the ordinal position of SNPs rather than the position of SNPs in the reference genome assembly.

Homozygosity in whole genome sequencing data can result from deletions, for which the affected region would have half the read depth of the genome-wide average, or copy-neutral losses of heterozygosity (which result rarely from meiotic recombination and central fusion parthenogenesis (Kronauer et al. 2012; Rabeling and Kronauer 2012; Oxley et al. 2014; Lacy et al. 2024)). To determine which of these processes caused homozygosity in diploid males, we identified runs of homozygosity using “bcftools roh” (Narasimhan et al. 2016), obtained read depths at all SNPs that passed filtering within these regions using “samtools depth –aa” (Danecek et al. 2021) and then normalized by dividing by the genome-wide median read depth. For samples without losses of heterozygosity on chromosome 4, we obtained normalized read depth at all SNPs that passed filtering on chromosome 4.

Homology Between CSD Loci Across Species

To determine whether our mapped CSD locus bore homology to those identified in other ant species, we used OrthoFinder (Emms and Kelly 2019) to identify orthologs of the genes present in the two CSD QTL in *V. emeryi* (Miyakawa and Mikheyev 2015). The locus identified in *L. humile* (Pan et al. 2024) is homologous to a subset of one of the *V. emeryi* QTL. In this process, we noticed that one of the genes found within our mapped CSD locus (*cysteine-rich with EGF-like domain protein 2, or CRELD2*) was also present in a second copy in a contig that was erroneously duplicated during the genome assembly process (Chr4:21571-39095). We masked this contig and realigned sequencing data for all subsequent analyses.

Balancing Selection

Because CSD loci are expected to evolve under negative frequency-dependent (balancing) selection, many different alleles should segregate in populations. To assess this, we used scikit-allel (Miles et al. 2019) to calculate nucleotide diversity across six different clonal lines and reported nucleotide diversity in sliding windows.

From inspecting DNA sequencing alignments to the reference genome, it became clear that the high genetic variation within the mapped CSD locus likely resulted from several highly differentiated haplotypes in this region. To investigate this further, we sequenced whole genomes of haploid males from clonal lines other than A and B. We assembled the genomes of each haploid male de novo using SPAdes (Bankevich et al. 2012; Prjibelski et al. 2020), found the contig bearing homology to the mapped CSD locus in the reference genome assembly using blastn (Camacho et al. 2009), manually scaffolded contigs as needed, and pairwise aligned each alternate allele to the reference genome using nucmer and mummer (Kurtz et al. 2004) to identify differences between alleles.

RNAseq for Annotation Improvements

To avoid missing any previously unannotated genes in the mapped CSD locus region, we performed RNA sequencing following various approaches. To start, we extracted total RNA from multiple samples, including separate pooled tissue from several different life stages (eggs, young larvae, fourth instar larvae, prepupae, pupae, and adults), different adult tissues (antennae, heads, thoraces, legs, abdomens), and individual male and female pupae using a TriZol RNA extraction method followed by precipitation in isopropanol at -20C overnight. We used the different life stage extractions for long read RNA-seq, rRNA depletion RNA-seq, and small RNA-seq. We used the different adult tissues for long read RNA-seq, and the individual male and female pupae for mRNA-seq. Long read RNA-seq was performed on a PacBio Sequel II, and all other sequencing was performed on an Illumina NovaSeq 6000. For rRNA depletion, we used Illumina's TruSeq Stranded Total RNA RiboZero kit. For small RNA-seq, we prepared libraries using Perkin Elmer's NEXTFLEX Small RNA-Seq Kit v3 for Illumina Platforms.

Annotation Improvements – Long Read Transcriptome Assembly

The GenBank reference genome (GCA_003672135) for *O. biroi* (Obir_v5.4) was retrieved from NCBI in FASTA format alongside the GenBank (GCA_003672135) and RefSeq (GCF_003672135.1) transcriptomes in GTF format. Additionally, a curated RefSeq annotation was retrieved from Zenodo (https://doi.org/10.5281/zenodo.10079884) in GTF format. RefSeq assemblies were chromosome name-mapped to GenBank chromosome names using the Bioconductor GenomeInfoDb package (Arora et al. 2024) and the chromosome name maps available in NCBI.

The PacBio toolset's pbmm2 (https://github.com/PacificBiosciences/pbmm2), a PacBio wrapper for Minimap2 (Li 2018), was used for indexing and alignment of pooled full-length, non-concatemer long reads to the GenBank reference genome using the parameters "--preset ISOSEQ -j 8". Following the generation of aligned long reads in BAM format, Stringtie2 (Kovaka et al. 2019) was used with parameters "-L" for long read transcriptome assembly, and the curated RefSeq annotation was supplied as a guide for assembly using the "-G" parameter. The resulting gene models were exported for further analysis and integration as a gzipped GTF file.

Annotation Improvements – Short Read Transcriptome Assembly

For Illumina RNA-seq data, FastQ quality control was performed using Rfastp (Wang and Carroll 2024). Indexing and alignment of short reads to the GenBank genome was performed using Hisat2 (Kim et al. 2019) with default parameters, and samtools was used to generate sorted and indexed BAM files for further analysis. Stringtie2 was then used to assemble transcriptomes using the default parameters. Gene models were then combined across replicates using Stringtie2's merge function with parameters "-F 0.5 -T 2 -f 0.1" and were exported for further analysis and integration as a gzipped GTF file.

Annotation Improvements – Integrating Short and Long Read Transcriptomes with the RefSeq Assembly.

Integration of long read and short read generated gene models with the established and curated RefSeq transcriptome was performed using custom scripts in R available from GitHub (https://github.com/RockefellerUniversity/Obiroi_GeneModels_2025) and described here.

First, conflicting gene models where single long read genes map to multiple RefSeq genes and vice versa were identified. As generated long read gene models are supported by multiple full-length, non-concatemer long read alignments, attention was focused on the 720 instances of multiple long read generated genes overlapping single RefSeq genes.

Single exon long read genes were removed as these are often found to be artifacts from PacBio sequencing. Following this, long read gene models that split RefSeq gene models but matched with GenBank gene models were accepted with the corresponding RefSeq models rejected.

Next, RefSeq gene models that conflicted within themselves were removed. This includes genes that were overlapping in their exons but marked as separate genes.

After tidying RefSeq gene models, long read gene models that split RefSeq gene models but matched with short read gene models were accepted, with the corresponding RefSeq models rejected. After this gene model polishing, only 132 gene models were left where multiple long read generated genes overlapped single RefSeq genes.

To potentially rescue internally primed transcripts caused by long stretches of genomic polyAs, long read transcripts were assessed for polyA nucleotide content downstream from their endpoints, and these distributions were visualized to identify cut-offs for putative internally primed transcripts. Transcripts with at least 70 percent as 10bp downstream of the last exon or with a polyA stretch > 4 in 10bp of the last exons were labeled as putative internally primed transcripts. For putative internally primed transcripts, which also split RefSeq genes, short read gene models were used to guide the rescued models when both short read and RefSeq agreed on a single gene model.

Finally, for the remaining long read gene models that split RefSeq gene models, gaps between long read gene models were assessed in the short read and RefSeq gene models. If agreement in exon and junction structure was found between short read and RefSeq gene models, the long read gene models were rejected. This left a final 27 genes where long read gene models split RefSeq gene models, and these were manually assessed in IGV alongside short read and long read alignments to reject or accept.

With a final set of long read gene models, all RefSeq and short read transcripts that fell within single genes were then added to the long read transcriptome to create an integrated transcriptome. Duplicate transcripts were then removed from the integrated transcriptome, and transcripts and genes were annotated for their sources (long read, short read, RefSeq) and any overlap with the established RefSeq, GenBank, and Ensembl gene models included. Following the creation of the integrated gene models, SQANTI3 (Pardo-Palacios et al. 2024) was used to provide a final corrected set of gene models as well as additional QC, as described below.

Assessment of Annotation Quality and Identifying Novel Genes, Transcripts and Exons

SQANTI3 was run with default parameters on the integrated transcriptome against the curated RefSeq and GenBank gene models to generate QC and descriptive statistics on novel and non-canonical splicing events. Gffcompare software (Pertea and Pertea 2020) was run against the curated RefSeq and GenBank gene models to provide statistics on novel exons, transcripts, and genes.

BUSCO (Simão et al. 2015) was run on the integrated, curated RefSeq and GenBank gene models with the parameters “-l insecta_odb10 -m tran” to provide measures of transcriptome completeness based on a set of core gene models. All these metrics were

summarized into tables using custom R scripts available on GitHub (https://github.com/RockefellerUniversity/Obiroi_GeneModels_2025).

Annotation of CDS and Functional Annotation of Genes

Annotation of coding regions was performed using TransDecoder (Haas 2023) to add CDS information to our integrated gene models GTF. TransDecoder was run with default parameters except for providing the "--single_best_only" to the TransDecoder.Predict function to get only one prediction per transcript. CDS information was merged with SQANTI3's GeneMark CDS annotation for a final set of CDSs. With this coding information, gene models were read into R, translated into protein sequences using the BioStrings package, and protein per transcript sequences exported as FASTA format for use in downstream tools. For functional annotation, both eggNOG (Huerta-Cepas et al. 2019) and InterProScan (Jones et al. 2014) were used using default parameters with our protein sequences FASTA to provide predictions for functional annotations of genes including their putative GO, KEGG, PFAM and Panther categories. Annotations were summarized to gene levels from transcript level annotation using R.

Annotation of lncRNAs

FEELnc (Wucher et al. 2017) was used to identify putative lncRNAs within our final integrated gene models. First, the FEELnc_filter.pl script was run with default parameters, and then the FEELnc_codpot.pl script was run with parameters "--mode=shuffle" to generate our putative set of lncRNAs for annotation of our integrated gene model GTF.

Annotation of miRNAs

Small RNA-seq reads were adaptor clipped for a “TGGAATTCTCGGGTGCCAAGG” adapter using CLIPflexR (Rozen-Gagnon et al. 2021) and aligned to the genome using miRDeep2's mapper.pl script with parameters “-c -j -m -l 18” (Friedländer et al. 2012). Aligned data were then processed using miRDeep2's miRDeep2.pl script with a FASTA of known *Drosophila* miRNAs retrieved from miRBase (Kozomara et al. 2019). The resulting annotated miRNA genomic locations were merged across samples and added into our integrated gene model GTF.

Annotation of piRNAs

For the identification of piRNA clusters, small read RNA-seq was processed using the pipeline workflow described within the Pilfer (Ray and Pandey 2018) toolkit to generate piRNA cluster BED files per sample. The sample piRNA clusters were then merged across samples to produce a final piRNA cluster set, which was then included in our integrated gene model GTF.

Annotation of tRNAs, snRNAs, snoRNAs and rRNAs

Additional annotations for tRNAs, snRNAs, snoRNAs and rRNAs were mapped from RefSeq annotation onto our integrated gene model GTF.

Sex-specific Gene Expression

To identify genes with sex-specific expression, we aligned Illumina mRNA-seq reads from three female and three male pupae using STAR (v2.5.0a) (Dobin et al. 2013). We counted and assigned reads to genes and transcripts using RSEM (v1.2.28) (Li and Dewey 2011). We analyzed differential gene expression using DESeq2 (Love et al. 2014), filtering out genes shorter than ten bases or with less than ten counts across all six samples. We hierarchically

clustered the differentially expressed genes by Euclidian distance using the R package pheatmap (v1.0.12) (Kolde 2024). We used DEXSeq (v1.50.0) (Anders et al. 2012) to identify differential exon usage.

Data and Software Availability

All DNA and RNA sequencing data are publicly available at the National Center for Biotechnology Information Sequence Read Archive under accession number PRJNA1075055. All other data are available in the article and Supporting Information. All code is available on GitHub (genome annotation: https://github.com/RockefellerUniversity/Obiroi_GeneModels_2025; differential gene expression and exon usage: https://github.com/jina-leemon/CSD_biroi_male-female-RNA-seq; all other code and analyses: https://github.com/kipdlacy/SexDetermination_Obiroi).

Contributions

K.D.L., J.L., and D.J.C.K. designed research; J.L. analyzed data and visualized results for gene-expression and exon usage analyses; K.R.G., W.W., and T.C. improved the genome annotation with input from K.D.L.; K.D.L. performed research, analyzed data, and visualized results for all other parts of the manuscript; K.D.L. and D.J.C.K. wrote the manuscript; D.J.C.K. supervised the project.

Competing interests

The authors declare no competing interests.

Acknowledgements

We thank Stephany Valdés-Rodríguez, Leonora Olivos-Cisneros, Alejandra Hurtado-Giraldo, Sean McKenzie, and other members of the Kronauer laboratory for helping us obtain male ants and for valuable discussions; Jen Balacco and the Rockefeller University Reference Genome Resource Center for long read RNA sequencing; the Weill Cornell Genomics Resources Core Facility for Small RNA sequencing; Connie Zhao and the Rockefeller University Genomics Resource Center for rRNA-depletion sequencing and all other Illumina sequencing; Tom Kay, Matteo Rossi, and Yukina Chiba for valuable comments on an earlier version of the manuscript. This work was supported by a Gabrielle H. Reem and Herbert J. Kayden Early-Career Innovation Award and the National Institute of General Medical Sciences of the National Institutes of Health under award no. R35GM127007, both to D.J.C.K. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This work was also supported by the Howard Hughes Medical Institute, where D.J.C.K. is an Investigator. This is Clonal Raider Ant Project paper number 36.

References

Anders S, Reyes A, Huber W. 2012. Detecting differential usage of exons from RNA-seq data. *Genome Research* 22:2008–2017.

Arora S, Morgan M, Carlson M, Pagès H. 2024. GenomeInfoDb: Utilities for manipulating chromosome names, including modifying them to follow a particular naming style.

Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman TL, Hahn MW, Kitano J, Mayrose I, Ming R, et al. 2014. Sex determination: why so many ways of doing it? *PLoS Biology* 12:e1001899.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 19:455–477.

Beye M, Hasselmann M, Fondrk MK, Page RE, Omholt SW. 2003. The gene *csd* is the primary signal for sexual development in the honeybee and encodes an SR-type protein. *Cell* 114:419–429.

Beye M, Seelmann C, Gempe T, Hasselmann M, Vekemans X, Fondrk MK, Page RE. 2013. Gradual molecular evolution of a sex determination switch through incomplete penetrance of femaleness. *Current Biology* 23:2559–2564.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.

Bopp D;, Saccone G;, Beye M. 2014. Sex determination in insects: variations on a common theme. *Sexual Development* 8:20–28.

Borowiec ML, Rabeling C, Brady SG, Fisher BL, Schultz TR, Ward PS. 2019. Compositional heterogeneity and outgroup choice influence the internal phylogeny of the ants. *Molecular Phylogenetics and Evolution* 134:111–121.

Borowiec ML, Zhang YM, Neves K, Ramalho MO, Fisher BL, Lucky A, Moreau CS. 2025. Evaluating UCE data adequacy and integrating uncertainty in a comprehensive phylogeny of ants. *Systematic Biology*:syaf001.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:1–9.

Chodroff RA, Goodstadt L, Sirey TM, Oliver PL, Davies KE, Green ED, Molnár Z, Ponting CP. 2010. Long noncoding RNA genes: conservation of sequence and brain expression among diverse amniotes. *Genome Biology* 11:R72.

Cook JM, Crozier RH. 1995. Sex determination and population biology in the Hymenoptera. *Trends in Ecology & Evolution* 10:281–286.

Crozier RH. 1971. Heterozygosity and sex determination in haplo-diploidy. *The American Naturalist* 105:399–412.

Dale RK, Pedersen BS, Quinlan AR. 2011. Pybedtools: a flexible Python library for manipulating genomic datasets and annotations. *Bioinformatics* 27:3423–3424.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience* 10:giab008.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21.

Duncan IW. 2002. Transvection effects in *Drosophila*. *Annual Review of Genetics* 36:521–556.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* 20:238.

Friedländer MR, Mackowiak SD, Li N, Chen W, Rajewsky N. 2012. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Research* 40:37–52.

Gempe T, Beye M. 2011. Function and evolution of sex determination mechanisms, genes and pathways in insects. *BioEssays* 33:52–60.

Gempe T, Hasselmann M, Schiøtt M, Hause G, Otte M, Beye M. 2009. Sex determination in honeybees: Two separate mechanisms induce and maintain the female pathway. *PLoS Biology* 7:e1000222.

Haas B. 2023. TransDecoder (find coding regions within transcripts). Available from: https://github.com/TransDecoder/TransDecoder

Hasselmann M, Beye M. 2004. Signatures of selection among sex-determining alleles of the honey bee. *Proceedings of the National Academy of Sciences* 101:4888–4893.

Hasselmann M, Gempe T, Schiøtt M, Nunes-Silva CG, Otte M, Beye M. 2008. Evidence for the evolutionary nascence of a novel sex determination pathway in honeybees. *Nature* 454:519–522.

Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, et al. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research* 47:D309–D314.

Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30:1236–1240.

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology* 37:907–915.

Koch V, Nissen I, Schmitt BD, Beye M. 2014. Independent Evolutionary Origin of fem Paralogous Genes and Complementary Sex Determination in Hymenopteran Insects. *PLOS ONE* 9:e91883.

Kolde R. 2024. pheatmap: Pretty Heatmaps. Available from: https://github.com/raivokolde/pheatmap

Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M. 2019. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biology* 20:278.

Kozomara A, Birgaoanu M, Griffiths-Jones S. 2019. miRBase: from microRNA sequences to function. *Nucleic Acids Research* 47:D155–D162.

Kronauer DJC, Pierce NE, Keller L. 2012. Asexual reproduction in introduced and native populations of the ant *Cerapachys biroi*. *Molecular Ecology* 21:5221–5235.

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. *Genome Biology* 5:R12.

Lacy KD, Hart T, Kronauer DJC. 2024. Co-inheritance of recombined chromatids maintains heterozygosity in a parthenogenetic ant. *Nature Ecology & Evolution* 8:1522–1533.

Laidlaw HH Jr, Gomes FP, Kerr WE. 1956. Estimation of the number of lethal alleles in a panmictic population of *Apis mellifera L*. *Genetics* 41:179–188.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Available from: http://arxiv.org/abs/1303.3997

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15:550.

Ma W-J, Schwander T. 2017. Patterns and mechanisms in instances of endosymbiont□induced parthenogenesis. *Journal of Evolutionary Biology* 30:868–888.

Matthey-Doret C, van der Kooi CJ, Jeffries DL, Bast J, Dennis AB, Vorburger C, Schwander T. 2019. Mapping of multiple complementary sex determination loci in a parasitoid wasp. *Genome Biology and Evolution* 11:2954–2962.

McKenzie SK, Kronauer DJC. 2018. The genomic architecture and molecular evolution of ant odorant receptors. *Genome Research* 28:1757–1765.

Miles A, Ralph P, Rae S, Pisupati R. 2019. cggh/scikit-allel: v1.2.1. Available from: https://zenodo.org/records/3238280

Miyakawa MO, Mikheyev AS. 2015. QTL mapping of sex determination loci supports an ancient pathway in ants and honey bees. *PLoS Genetics* 11:e1005656.

Miyakawa MO, Miyakawa H. 2023. *Transformer* gene regulates feminization under two complementary sex determination loci in the ant, *Vollenhovia emeryi*. *Insect Biochemistry and Molecular Biology* 156:103938.

Miyakawa MO, Tsuchida K, Miyakawa H. 2018. The doublesex gene integrates multi-locus complementary sex determination signals in the Japanese ant, *Vollenhovia emeryi*. *Insect Biochemistry and Molecular Biology* 94:42–49.

Narasimhan V, Danecek P, Scally A, Xue Y, Tyler-Smith C, Durbin R. 2016. BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics* 32:1749–1751.

Normark BB. 2003. The evolution of alternative genetic systems in insects. *Annual Review of Entomology* 48:397–423.

Otte M, Netschitailo O, Weidtkamp-Peters S, Seidel CAM, Beye M. 2023. Recognition of polymorphic Csd proteins determines sex in the honeybee. *Science Advances* 9:eadg4239.

Oxley PR, Ji L, Fetter-Pruneda I, McKenzie SK, Li C, Hu H, Zhang G, Kronauer DJC. 2014. The genome of the clonal raider ant *Cerapachys biroi*. *Current Biology* 24:451–458.

Page RE Jr. 1980. The evolution of multiple mating behavior by honey bee queens (*Apis mellifera L.*). *Genetics* 96:263–273.

Pan Q, Darras H, Keller L. 2024. LncRNA gene ANTSR coordinates complementary sex determination in the Argentine ant. *Science Advances* 10:eadp1532.

Pardo-Palacios FJ, Arzalluz-Luque A, Kondratova L, Salguero P, Mestre-Tomás J, Amorín R, Estevan-Morió E, Liu T, Nanni A, McIntyre L, et al. 2024. SQANTI3: curation of long-read transcriptomes for accurate identification of known and novel isoforms. *Nature Methods* 21:793–797.

Pertea G, Pertea M. 2020. GFF Utilities: GffRead and GffCompare. Available from: https://f1000research.com/articles/9-304

Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Auwera GAV der, Kling DE, Gauthier LD, Levy-Moonshine A, Roazen D, et al. 2018. Scaling accurate genetic variant discovery to tens of thousands of samples. Available from: https://www.biorxiv.org/content/10.1101/201178v3

Privman E, Wurm Y, Keller L. 2013. Duplication and concerted evolution in a master sex determiner under balancing selection. *Proceedings of the Royal Society B: Biological Sciences* 280:20122968.

Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. 2020. Using SPAdes De Novo Assembler. *Current Protocols in Bioinformatics* 70:e102.

Quinn JJ, Zhang QC, Georgiev P, Ilik IA, Akhtar A, Chang HY. 2016. Rapid evolutionary turnover underlies conserved lncRNA–genome interactions. *Genes & Development* 30:191–207.

Rabeling C, Kronauer DJC. 2012. Thelytokous parthenogenesis in eusocial Hymenoptera. *Annual Review of Entomology* 58:273–292.

Ray R, Pandey P. 2018. piRNA analysis framework from small RNA-Seq data by a novel cluster prediction tool - PILFER. *Genomics* 110:355–365.

Rozen-Gagnon K, Gu M, Luna JM, Luo J-D, Yi S, Novack S, Jacobson E, Wang W, Paul MR, Scheel TKH, et al. 2021. Argonaute-CLIP delineates versatile, functional RNAi networks in *Aedes aegypti*, a major vector of human viruses. *Cell Host & Microbe* 29:834–848.

Sawanth SK, Gopinath G, Sambrani N, Arunkumar KP. 2016. The autoregulatory loop: A common mechanism of regulation of key sex determining genes in insects. *J Biosci* 41:283–294.

Schmieder S, Colinet D, Poirié M. 2012. Tracing back the nascence of a new sex-determination pathway to the ancestor of bees and ants. *Nature Communications* 3:1–7.

Seiler J, Beye M. 2024. Honeybees' novel complementary sex-determining system: function and origin. *Trends in Genetics* 40:969–981.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212.

Suomalainen E, Saura A, Lokki J. 1987. Cytology and Evolution in Parthenogenesis. CRC Press

Trible W, Chandra V, Lacy KD, Limón G, McKenzie SK, Olivos-Cisneros L, Arsenault SV, Kronauer DJC. 2023. A caste differentiation mutant elucidates the evolution of socially parasitic ants. *Current Biology* 33:1047–1058.

Trible W, Mckenzie SK, Kronauer DJC. 2020. Globally invasive populations of the clonal raider ant are derived from Bangladesh. *Biology Letters* 16:20200105.

Ulitsky I, Shkumatava A, Jan CH, Sive H, Bartel DP. 2011. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* 147:1537–1550.

Verhulst EC, Beukeboom LW, van de Zande L. 2010. Maternal control of haplodiploid sex determination in the wasp *Nasonia*. *Science* 328:620–623.

Vorburger C. 2013. Thelytoky and Sex Determination in the Hymenoptera: Mutual Constraints. *Sexual Development* 8:50–58.

Wang W, Carroll T. 2024. Rfastp: An ultra-fast and all-in-one fastq preprocessor (quality control, adapter, low quality and polyX trimming) and UMI sequence parsing). Available from: 10.18129/B9.bioc.Rfastp

Wexler J, Delaney EK, Belles X, Schal C, Wada-Katsumata A, Amicucci MJ, Kopp A. 2019. Hemimetabolous insects elucidate the origin of sexual development via alternative splicing. *eLife* 8:e47490.

Whiting PW. 1933. Selective fertilization and sex-determination in Hymenoptera. *Science* 78:537–538.

Whiting PW. 1935. Genic balance, sex-determination and selective fertilization in Hymenoptera. *Proceedings of the American Philosophical Society* 75:517–520.

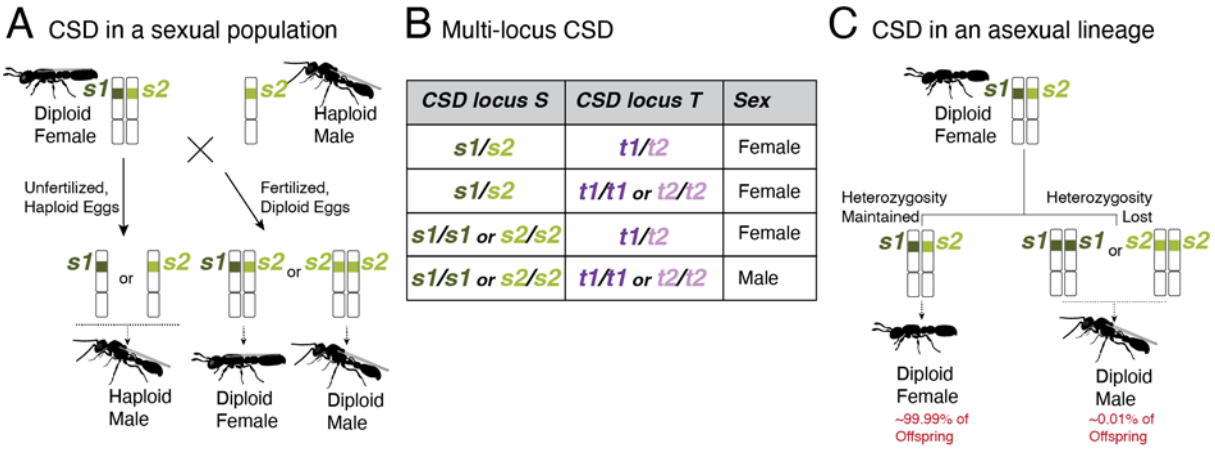
Whiting PW. 1943. Multiple alleles in complementary sex determination of *Habrobracon*. *Genetics* 28:365–382.

van Wilgenburg E, Driessen G, Beukeboom LW. 2006. Single locus complementary sex determination in Hymenoptera: an “unintelligent” design? *Frontiers in Zoology* 3:1.

Williams TM, Carroll SB. 2009. Genetic and molecular insights into the development and evolution of sexual dimorphism. *Nature Reviews Genetics* 10:797–804.

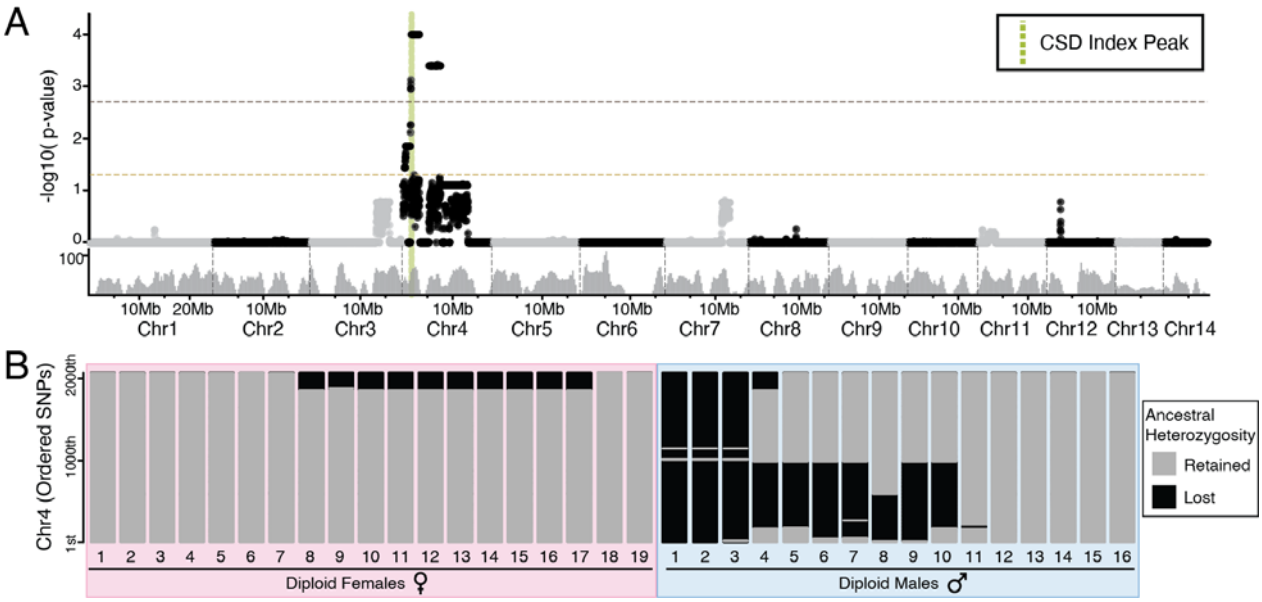
Wucher V, Legeai F, Hédan B, Rizk G, Lagoutte L, Leeb T, Jagannathan V, Cadieu E, David A, Lohi H, et al. 2017. FEELnc: a tool for long non-coding RNA annotation and its application to the dog transcriptome. *Nucleic Acids Research* 45:e57.

Zou Y, Geuverink E, Beukeboom LW, Verhulst EC, van de Zande L. 2020. A chimeric gene paternally instructs female sex determination in the haplodiploid wasp *Nasonia*. *Science* 370:1115–1118.



CSD locus S	CSD locus T	Sex
s1/s2	t1/t2	Female
s1/s2	t1/t1 or t2/t2	Female
s1/s1 or s2/s2	t1/t2	Female
s1/s1 or s2/s2	t1/t1 or t2/t2	Male

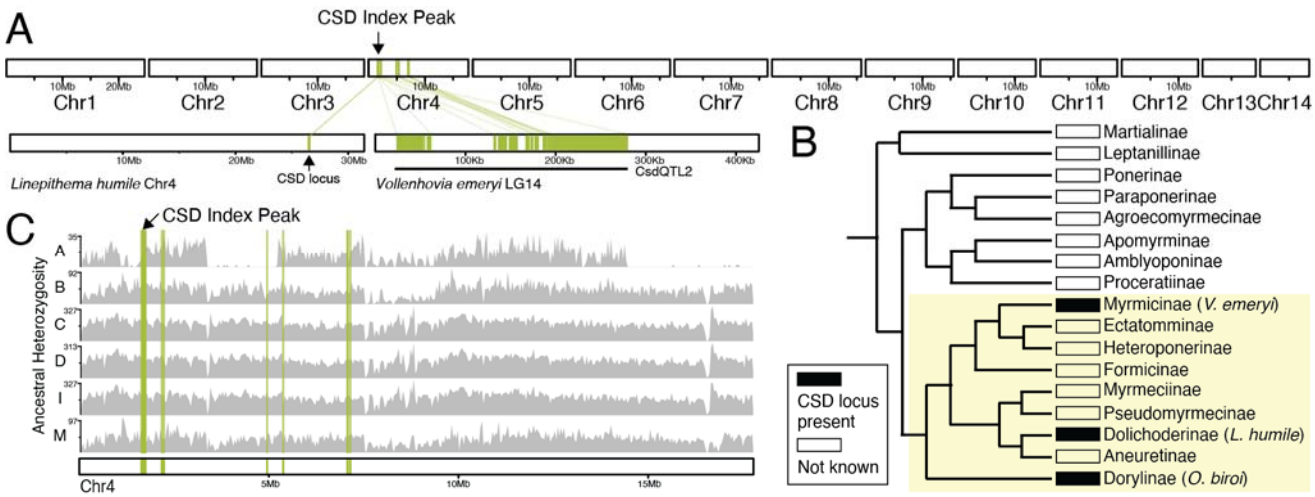
Fig. 1. Theoretical expectations for complementary sex determination (CSD). (A) Cartoon depicting a mating between a diploid female with two different alleles at a sex determination locus and a haploid male bearing one of these two alleles. Half of the sexually produced diploid offspring are expected to develop as males. **(B)** Table illustrating multi-locus CSD in diploids, in which heterozygosity of at least one sex determination locus is required for female development. By contrast, only homozygosity at all loci results in diploid male development. **(C)** Cartoon depicting how CSD might work in asexual species such as the clonal raider ant. If diploid males arise from losses of heterozygosity at sex loci, then homozygosity for either of the two alleles should trigger male development. Offspring proportions reflect empirical observations in the clonal raider ant (Kronauer et al. 2012; Oxley et al. 2014).



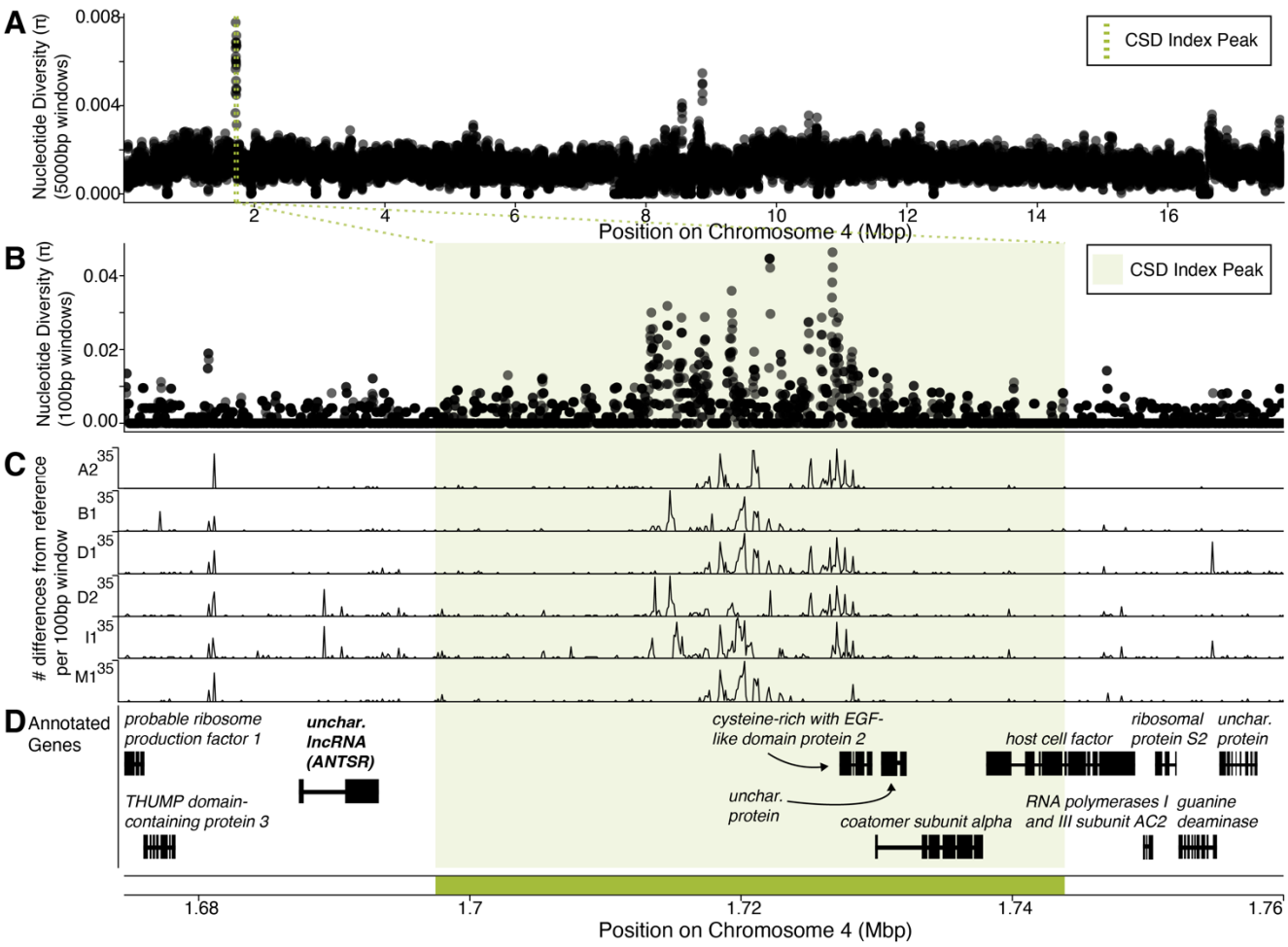
<u>Fig. 2.</u> Whole genome sequencing reveals a candidate sex determination locus on chromosome 4. (A) Karyoplot depicting the mean CSD index p-value in 50kb windows with a 15kb sliding interval. The CSD index peak is shown as a green dotted line. The significance threshold (p=0.05) and FDR-corrected significance threshold (p=0.002) are indicated by gold and brown lines, respectively. The grey histogram shows the number of ancestrally heterozygous SNPs in 300kb windows. **(B)** Stacked bar plots for all diploids used for mapping, with one horizontal line for each ordered putatively ancestrally heterozygous SNP on chromosome 4. For each sample, SNPs that retain ancestral heterozygosity are drawn as grey lines, whereas SNPs that have lost ancestral heterozygosity are drawn as black lines.

	Heterozygous	**Homozygous Allele 1**	**Homozygous Allele 2**
Diploid Females	19	0	0
Diploid Males	5	2	9

Table 1. Number of samples heterozygous and homozygous for each allele in the 46kb region on chromosome 4.



<u>Fig. 3.</u> The putative sex determination locus identified in *O. biroi* is conserved across formicoid ants. **(A)** Karyoplot depicting the 14 chromosomes in the *O. biroi* genome, chromosome 4 from the *L. humile* genome (Pan et al. 2024), and one contig from linkage group 14 of the *V. emeryi* reference genome (Miyakawa and Mikheyev 2015). Note that the plots from the different species are not drawn to scale. Homology to protein-coding genes identified within *V.emeryiCsdQTL2* is drawn in green. The location of the *O. biroi* CSD index peak is indicated. **(B)** A phylogeny of the ant subfamilies (adapted from (Borowiec et al. 2019)), with the presence of a sex determination locus homologous to the peak of our CSD index, or the absence of data shown. The yellow shaded background denotes the formicoid clade. **(C)** Karyoplot for *O. biroi* chromosome 4, depicting homology to *V.emeryiCsdQTL2* and ancestral heterozygosity for six different clonal lines (A, B, C, D, I, and M). Grey histograms depict the number of ancestrally heterozygous SNPs in 100kb windows.



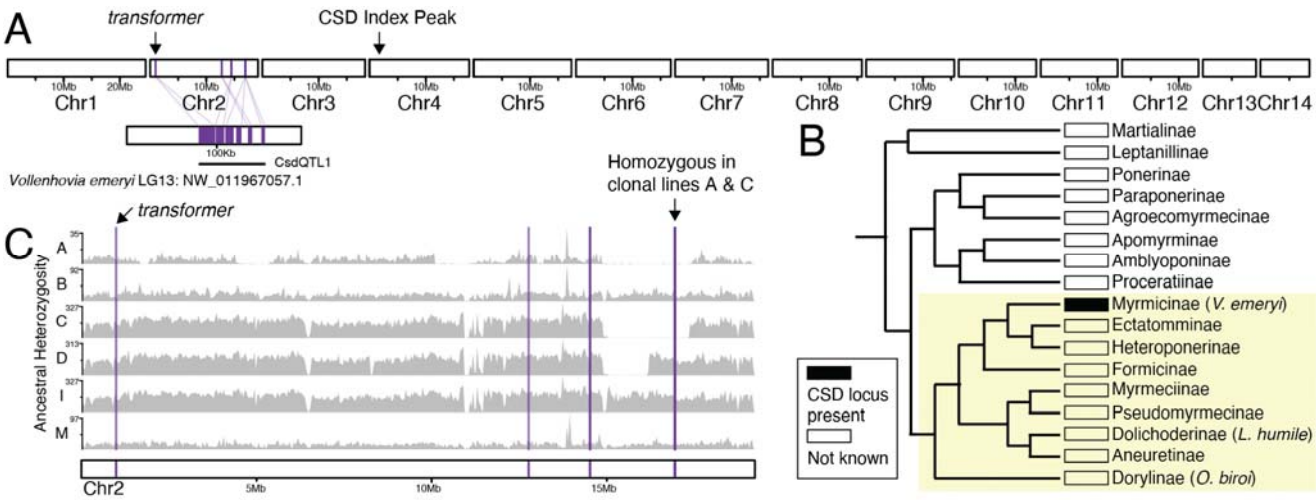
<u>Fig. 4.</u> The CSD index peak is characterized by high genetic diversity in a non-coding region. (A & B) Nucleotide diversity across the length of *O. biroi* chromosome 4 in 5kb windows (step size=1kb) **(A)** and across the vicinity of the CSD index peak in 100bp windows (step size=20bp) **(B)**. **(C)** The number of differences per 100bp window between each alternate *de novo* assembled allele and the reference genome allele. **(D)** Annotated genes in the vicinity of the CSD index peak. Black boxes depict exons and thin lines depict introns. The lncRNA *ANTSR* is indicated in bold. Arrows indicate the names of genes in close proximity. The CSD index peak is shown as a green dotted line in A, and with green shading in B-D.

Clonal Line	CSD Index Peak Haplotypes
A	A1 / A2
B	B1 / A1
C	? / ?
D	D1 / D2
I	I1 / ?
M	M1 / A1

Table 2. The haplotypes present at the CSD index peak for each studied *O. biroi* clonal line. Because no haploid males were sequenced from clonal line C, and only one haploid male was sequenced from clonal line I, the haplotypes for these two lines remain incompletely known. We denote the unidentified haplotypes with question marks.

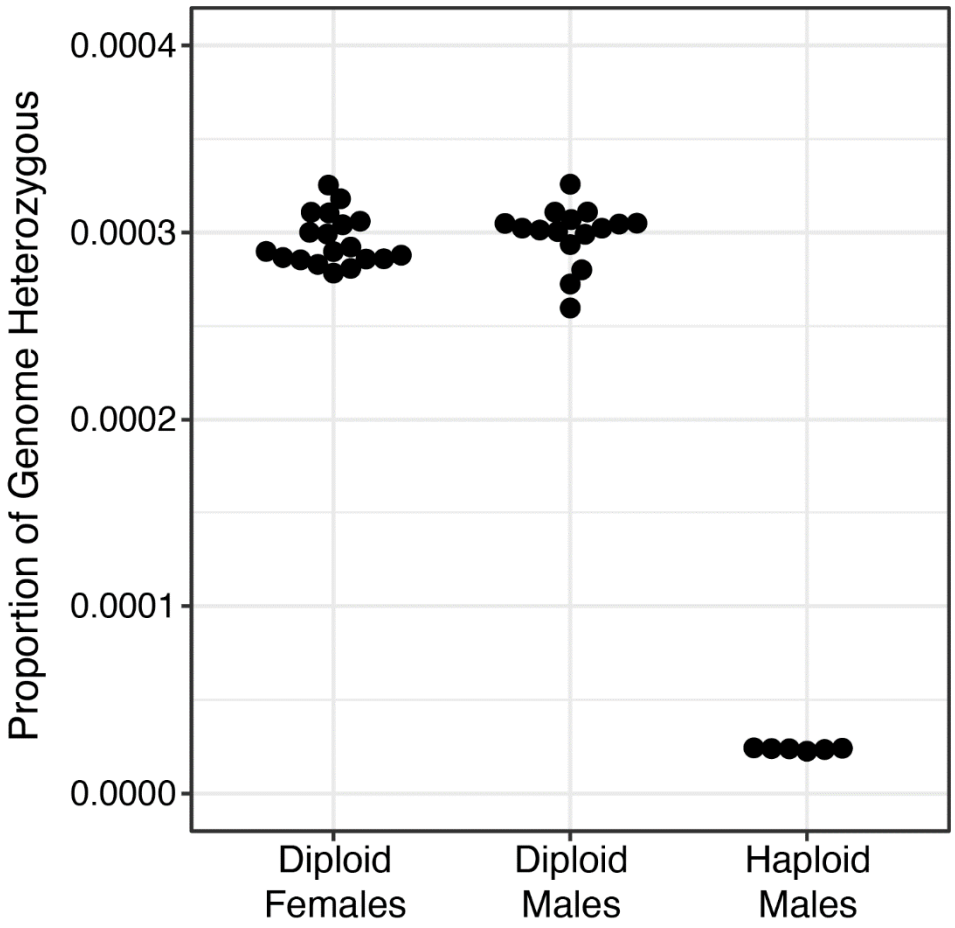
Clonal Line	**Haploid Males**	**Diploid Males**
A	92	22
B	78	0
C	6	0
D	4	0
I	1	0
M	2	0

Table 3. Number of haploid and diploid males sampled from different *O. biroi* clonal lines.

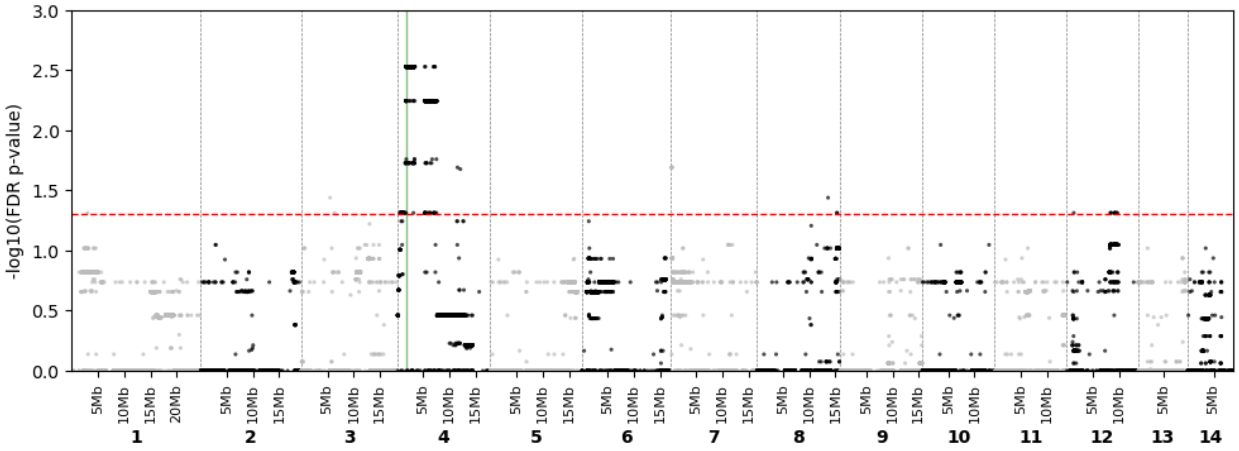


<u>Fig. 5.</u> Whether a second, *tra*-containing CSD QTL from *V. emeryi* is conserved across ants remains ambiguous. (A) Karyoplot depicting the 14 chromosomes in the *O. biroi* genome and one contig from linkage group 13 of the *V. emeryi* reference genome. Homology to protein-coding genes identified within *V.emeryiCsdQTL1* is drawn in purple. The locations of the *O. biroi* CSD index peak and the homolog of *transformer* are indicated. **(B)** A phylogeny of the ant subfamilies (adapted from (Borowiec et al. 2019)), with the presence or absence of this second putative sex determination locus, or the absence of data shown. The yellow shaded background denotes the formicoid clade. **(C)** Karyoplot for *O. biroi* chromosome 2, depicting homology to *V.emeryiCsdQTL1* and ancestral heterozygosity for each clonal line. Grey histograms depict the number of ancestrally heterozygous SNPs in 100kb windows. The region of homology that is ancestrally homozygous in clonal lines A and C is labeled.

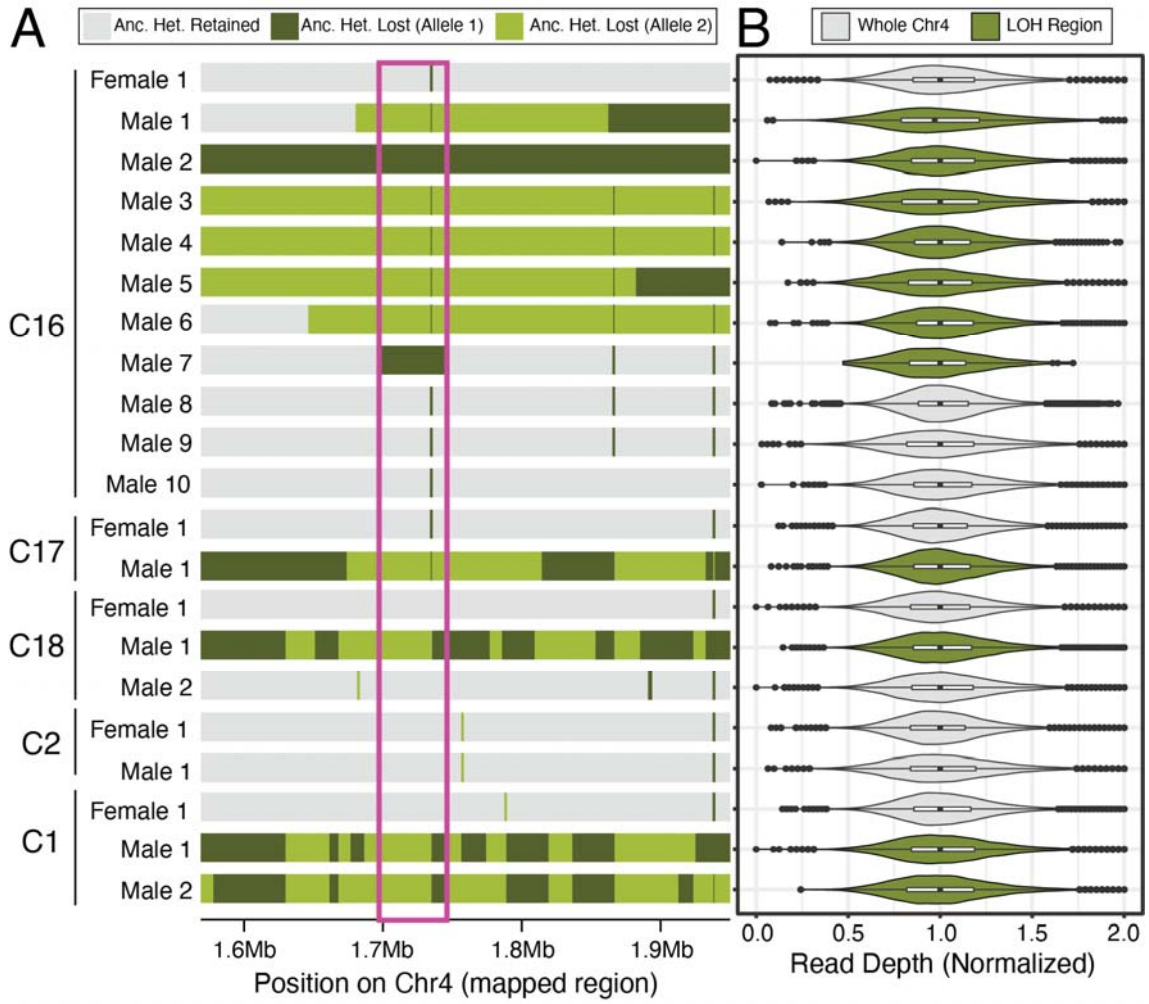
Supplementary Figures



Supplementary Fig. S1. Genome-wide heterozygosity levels. Scatterplot depicting the proportion of base pairs in the ~220Mb *O. biroi* genome that are heterozygous in each individual for which we conducted whole genome sequencing. Haploid males have no legitimate heterozygosity—the small elevation above zero is due to heterozygous genotype calls at erroneously assembled regions of the reference genome.

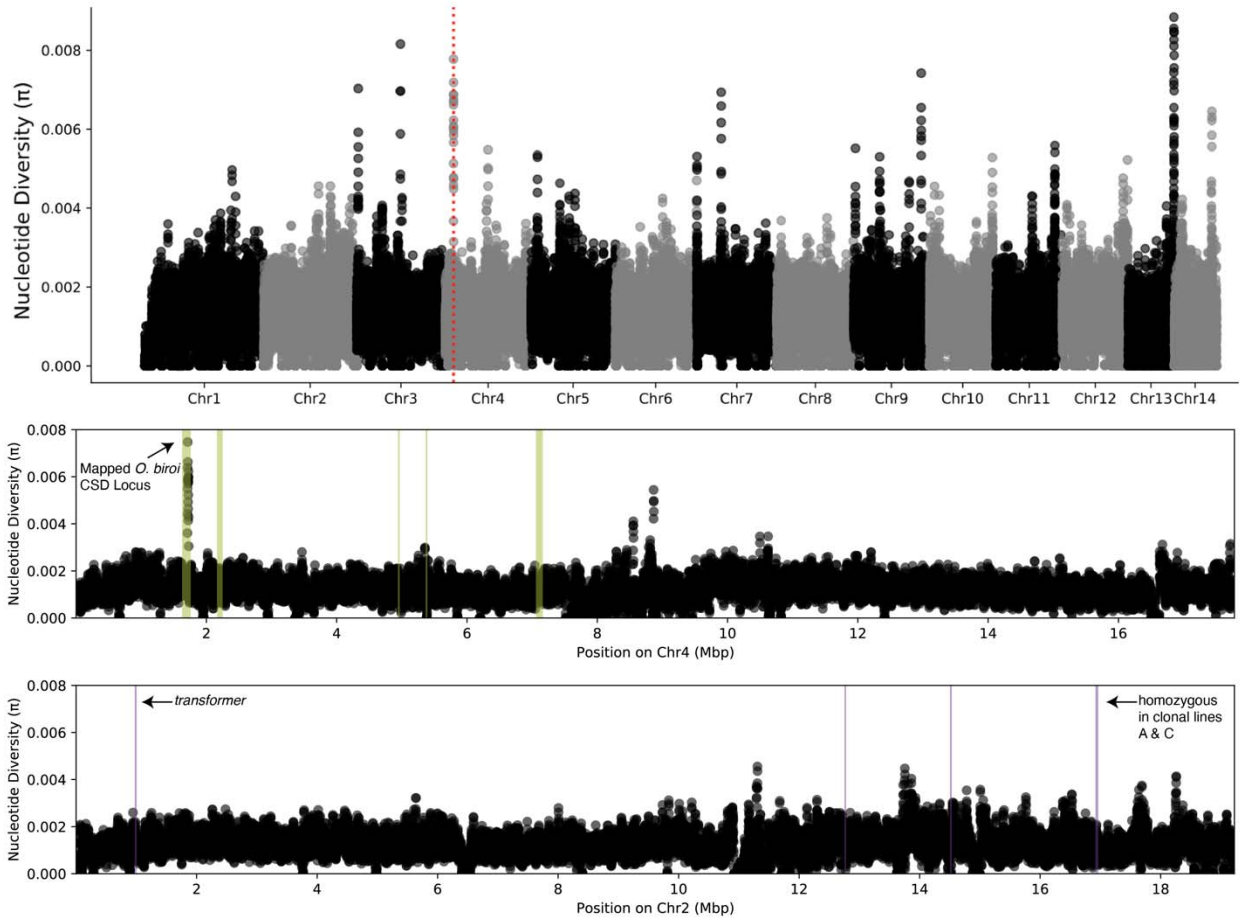


<u>Supplementary Fig. S2.</u> Homozygosity levels differ between males and females at the putative CSD locus. Manhattan plot depicting FDR-corrected p-values from Fisher's Exact Tests on whether homozygosity levels differ between males and females at each SNP. The red line indicates the significance threshold (p=0.05), and the vertical green line marks the peak of the CSD index.

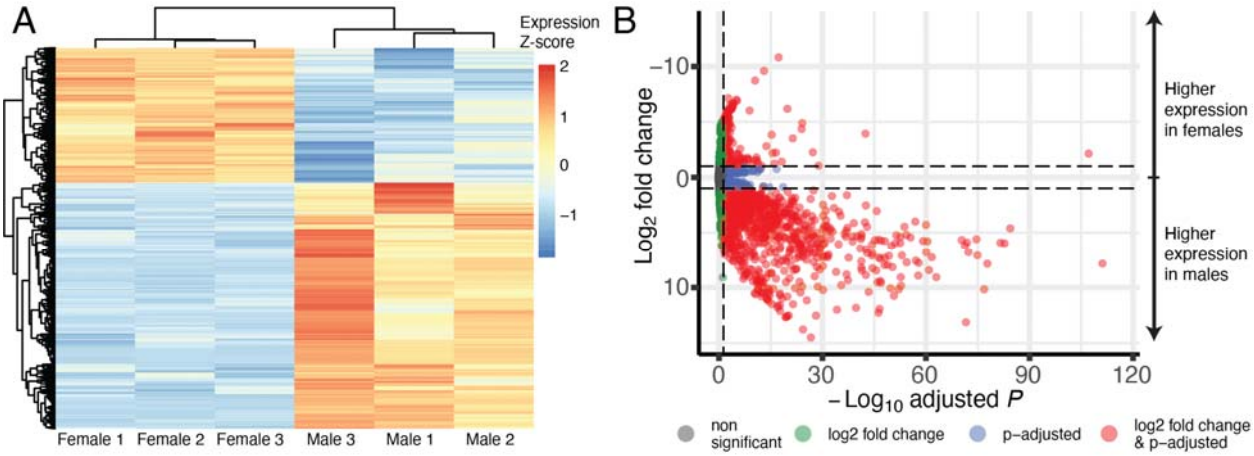


<u>Supplementary Fig. S3.</u> Diploid males result from copy-neutral losses of heterozygosity for either allele. (A) Karyoplot depicting losses (green) or retention (gray) of heterozygosity across the vicinity of the peak of the CSD index in diploid males and females. Codes at the left (C16, C17, etc.) indicate the stock colonies from which each ant was sampled. The magenta box surrounds the CSD index peak. Dark green indicates that the sample has the same allele / haplotype as the reference haploid male whereas light green indicates that the sample has the alternative allele. Switches between alleles in homozygous regions within samples (i.e., changes from light green to dark green or vice versa) represent recombination events that occurred in generations prior to the meiosis that produced each diploid male. Because the reference haploid

male was from colony C16, fewer generations have elapsed since the common ancestor of the reference haploid male and the diploid males from that colony, and thus fewer recombination events have accumulated. By contrast, more generations have elapsed since the common ancestor of the reference haploid male and the diploid males from colonies C17, C18, and C1, which is reflected in the higher number of recombination events that accumulated across generations. **(B)** Normalized read depth across either all SNPs that passed filtering on chromosome 4 (gray, for samples without losses of heterozygosity on chromosome 4) or all SNPs that passed filtering within that sample's loss of heterozygosity region (green). Box plots show median (center line), interquartile range (IQR) (box limits) and 1.5 x IQR (whiskers). Data points that fall outside 1.5 x IQR are shown individually. Violin plots show the kernel probability density, meaning that the proportion of the data located at an x-axis value is represented by the width of the outlined area.

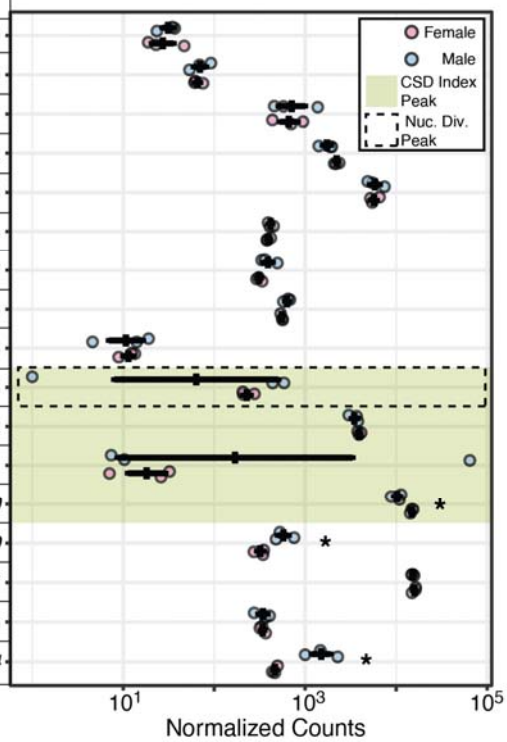


<u>Supplementary Fig. S4.</u> Genome-wide nucleotide diversity. Nucleotide diversity across the genome in 5kb windows with a 1kb sliding interval. In the genome-wide plot (top), the red line depicts the peak of the CSD index from genetic mapping. In the zoomed-in plots for chromosomes 4 (middle) and 2 (bottom), the green and purple vertical shading indicates homology to *V. emeryi* CSD QTL 2 and 1, respectively.



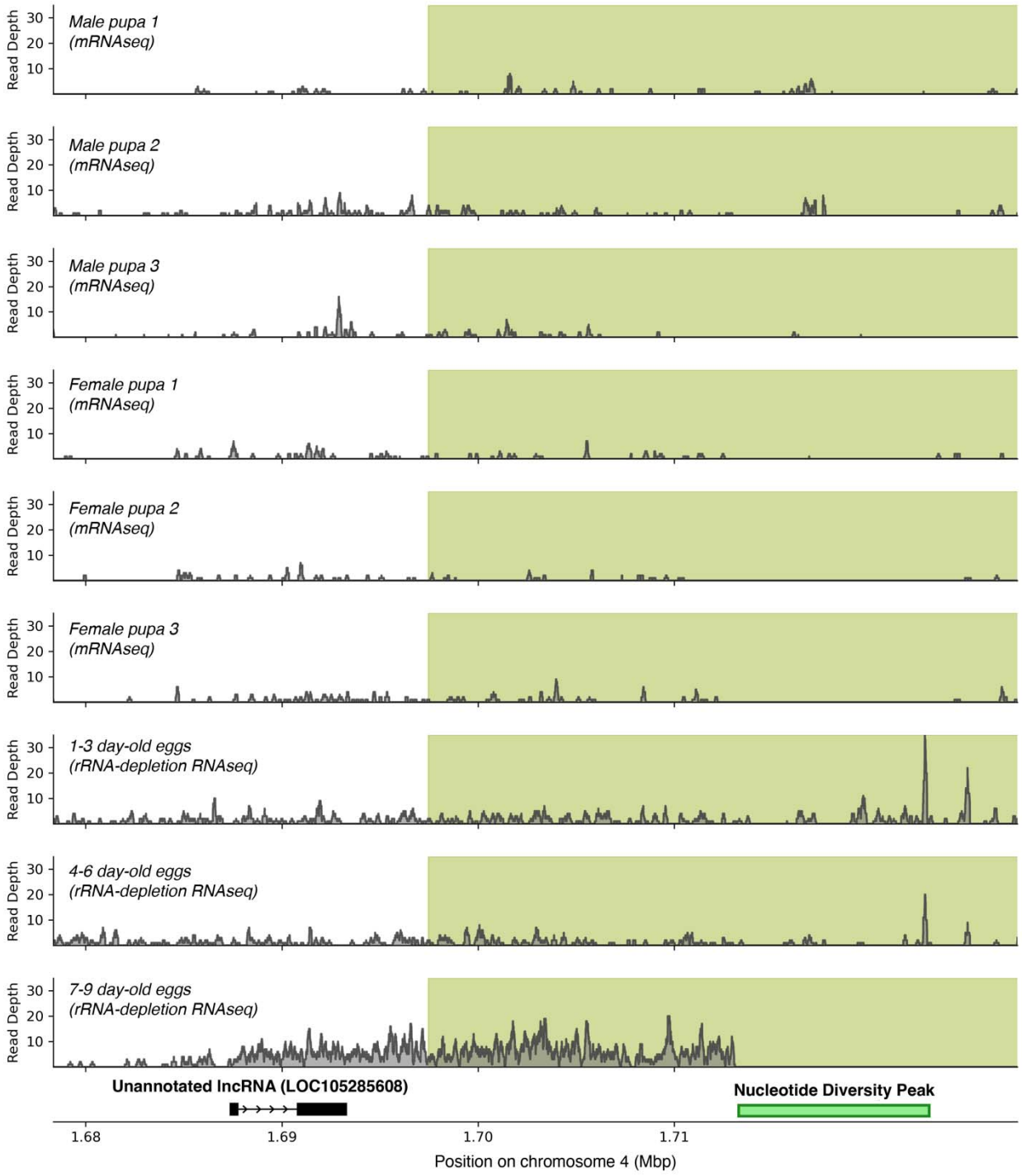
C

Pos. on chromosome 4	Gene Name	Gene ID
Chr4:1607254-1609373	unchar. protein	LOC105285599
Chr4:1609611-1612715	unchar. protein	LOC105285600
Chr4:1619377-1621011	paraneoplastic antigen Ma6E	LOC105285601
Chr4:1621560-1634471	ataxin-7	LOC105285602
Chr4:1637384-1672949	Kank	LOC105285603
Chr4:1673043-1674418	Bet1	LOC105285604
Chr4:1674523-1676002	prob. ribosome prod. factor 1	LOC105285607
Chr4:1675932-1678296	THUMP	LOC105285605
Chr4:1687395-1693282	unchar. lncRNA	LOC105285608
Chr4:1727256-1729679	Creld	LOC113561733
Chr4:1729888-1737844	alphaCOP	LOC105283852
Chr4:1730326-1732211	unchar. protein	LOC105283851
Chr4:1738076-1749057	**Host cell factor**	***LOC105283850***
Chr4:1749634-1750382	**l(2)37Cg**	***LOC105283849***
Chr4:1750527-1752116	ribosomal protein S2	LOC105283847
Chr4:1752262-1755091	guanine deaminase	LOC105283846
Chr4:1755270-1770406	**unchar. protein**	***LOC105283844***



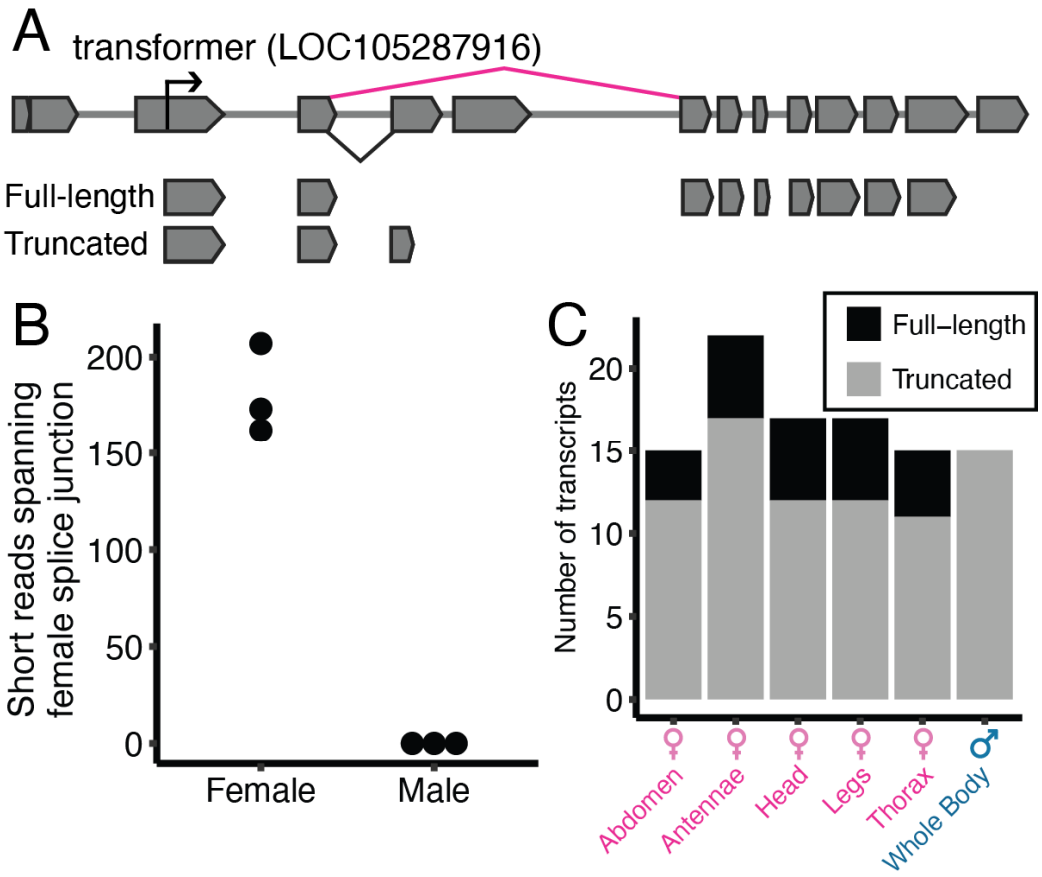
<u>Supplementary Fig. S5.</u> Sex-specific gene expression. (**A**) Heatmap of genes that were significantly differentially expressed between sexes (adjusted p-value < 0.05). Each row represents a gene, and each column represents a sample. The trees at the left and the top of the heat map depict hierarchical clustering of genes and samples, respectively. Z-scores of log-transformed expression levels are shown as a color gradient. (**B**) Volcano plot depicting differential gene expression between males and females. Each gene is represented by a dot,

which is colored gray if not differentially expressed, green if the expression difference is at least 2-fold (log□ ≥ 1), or blue if the expression difference is statistically significant (adjusted p-value < 0.05). Red indicates genes that are significantly differentially expressed and have at least a 2-fold difference in gene expression. (**C**) Expression levels of genes in the vicinity of the CSD index peak. Normalized counts are shown for each male and female sample, with error bars depicting the mean and SEM for each sex. The genomic coordinates, name, and ID are given for each gene. Genes with significantly different expression between sexes (adjusted p-value < 0.05) are bolded in the table and marked with an asterisk on the plot. The CSD index peak is shown in green, and the dashed box denotes the peak of nuclear diversity.



<u>Supplementary Fig. S6.</u> RNAseq reads aligned to the unannotated lncRNA flanking the putative CSD locus. RNAseq read depth in the vicinity of the unannotated lncRNA in nine different libraries. The peak of the CSD index is indicated in light green shading, but note that

this region extends further downstream. The nucleotide diversity peak is indicated by a horizontal green bar.



<u>Supplementary Fig. S7.</u> The *O. biroi* transformer homolog is sex-specifically spliced. The c-terminal region of *tra* exon four is spliced differently in females (pink) and males (blue), producing a truncated protein in males and a functional, full-length TRA protein in females. (**A**) Diagram of transformer splicing depicting full-length (functional) and truncated (non-functional) transcripts. The black arrow indicates the transcription start site, the pink line indicates the female-specific splice junction that allows the production of functional protein, and the black line indicates the splice junction that leads to a truncated protein. (**B**) The number of reads spanning the female splice junction required to produce functional protein from three female and three male short-read RNAseq libraries. (**C**) The number of full-length and truncated *tra* transcripts in six long read RNA sequencing (IsoSeq) libraries from different tissues and sexes.

Supplementary Tables

Primer Name	5.4 Scaffold	5.4 Position	Forward 5'-3'	Reverse 5'-3'	Genotyping Method
Obir Ploidy 12SV	Chr12	12801234	TAGATCGAGTGAGGTATCTT	AGCGTTAAGAGTAGATTAAG	Visualize differently sized alleles (i.e., different versions of a structural variant [SV]) using gel electrophoresis
Obir Ploidy 4RE	Chr4	5840887	CCACCCTCTCTCTCGTCTC	TGCCGATTCATATTTCGAGATAAAC	Separately digest products with the restriction enzymes (REs) EcoRI and BsmI and visualize genotypes via gel electrophoresis
Obir Ploidy 5RE	Chr5	1963279	GCGATAGAAGAGTGCTCGTATT	GCCAGTCAGTCTGTCATTCA	Separately digest products with the restriction enzymes (REs) EcoRI and AvaII and visualize genotypes via gel electrophoresis
Obir Ploidy 2SNP	Chr2	18082503	CACGTACAATGAGGGAACCTAA	TCGCGATCATGAAATCTAACCT	Sanger sequence and call SNPs
Obir Ploidy 12SNP	Chr12	8100152	GTCAGTGCGCTGTTCATTTC	AGATTATAGGGCGTTCCAATTCT	Sanger sequence and call SNPs
Obir Ploidy 14SNP	Chr14	6377499	CACGTGAACGCTACGATGAA	GGGTTGAGGATGTCTAAGGTAATG	Sanger sequence and call SNPs
Obir Ploidy 13SNPa	Chr13	6262370	CGTTACACTCCTGTGTCGAATC	GCACGTAGCATTCTCTCTTTCT	Sanger sequence and call SNPs
Obir Ploidy 13SNPb	Chr13	7173724	GCCATACATCCTGCACTCTTAC	CGACGAGCATGCGTCTAAA	Sanger sequence and call SNPs

Supplementary Table S1. Primers for ploidy assessment via heterozygosity. 5.4 refers to the reference genome version (GCA_003672135.1).

Sample name	Library prep method	Clonal line	Stock colony	Sample type	Mean read depth	BioProject
LineA_DiploidFemale_1	Nextera Flex	A	C16	Diploid Female	30.1	PRJNA923657
LineA_DiploidFemale_2	Nextera Flex	A	C16	Diploid Female	27.4	PRJNA923657
LineB_DiploidFemale_1	Nextera Flex	B	STC6	Diploid Female	28.3	PRJNA923657
LineA_DiploidFemale_6	Nextera Flex	A	C18	Diploid Female	30	PRJNA923657
LineA_DiploidFemale_5	Nextera Flex	A	C1	Diploid Female	41.5	PRJNA923657
LineA_DiploidFemale_3	Nextera Flex	A	C16	Diploid Female	26.6	PRJNA947942
LineA_DiploidFemale_4	Nextera Flex	A	C16	Diploid Female	24.1	PRJNA947942
LineB_DiploidFemale_2	Nextera Flex	B	STC6	Diploid Female	27	PRJNA947942
LineB_DiploidFemale_3	Nextera Flex	B	STC6	Diploid Female	47.1	PRJNA947942
LineA_HaploidMale_1	Nextera Flex	A	C16	Haploid Male	26.1	PRJNA947942
LineA_HaploidMale_2	Nextera Flex	A	C16	Haploid Male	26.1	PRJNA947942
LineA_HaploidMale_3	Nextera Flex	A	C16	Haploid Male	27.6	PRJNA947942
LineA_HaploidMale_4	Nextera Flex	A	C16	Haploid Male	25.8	PRJNA947942
LineB_HaploidMale_1	Nextera Flex	B	STC6	Haploid Male	26.3	PRJNA947942
LineB_HaploidMale_2	Nextera Flex	B	STC6	Haploid Male	27.6	PRJNA947942
LineB_HaploidMale_3	Nextera Flex	B	STC6	Haploid Male	31.7	PRJNA947942
LineB_DiploidFemale_5	Nextera Flex	B	STC12	Diploid Female	42.4	PRJNA947942
LineB_DiploidFemale_4	Nextera Flex	B	STC1	Diploid Female	31.7	PRJNA947942
LineA_HaploidMale_5	Nextera Flex	A	C1	Haploid Male	30.3	PRJNA947942
LineA_HaploidMale_6	Nextera Flex	A	C18	Haploid Male	25.3	PRJNA947942
LineB_HaploidMale_4	Nextera Flex	B	STC1	Haploid Male	31	PRJNA947942
LineB_HaploidMale_5	Nextera	B	STC12	Haploid	32.8	PRJNA947

	Flex			Male		942
DiploidMale_1	Nextera Flex	A	C16	Diploid Male	57.9	PRJNA107 5055
DiploidMale_2	Nextera Flex	A	C16	Diploid Male	32.7	PRJNA107 5055
DiploidMale_3	Nextera Flex	A	C16	Diploid Male	31.5	PRJNA107 5055
DiploidMale_4	Nextera Flex	A	C18	Diploid Male	38.3	PRJNA107 5055
DiploidMale_5	Nextera Flex	A	C17	Diploid Male	47.7	PRJNA107 5055
DiploidMale_6	Nextera Flex	A	C16	Diploid Male	28.4	PRJNA107 5055
DiploidMale_7	Nextera Flex	A	C16	Diploid Male	32.0	PRJNA107 5055
DiploidMale_8	Nextera Flex	A	C2	Diploid Male	30.2	PRJNA107 5055
DiploidMale_9	Nextera Flex	A	C16	Diploid Male	41.6	PRJNA107 5055
DiploidMale_10	Nextera Flex	A	C16	Diploid Male	28.1	PRJNA107 5055
DiploidMale_11	Nextera Flex	A	C18	Diploid Male	40.2	PRJNA107 5055
DiploidMale_12	Nextera Flex	A	C16	Diploid Male	38.5	PRJNA107 5055
DiploidMale_13	Nextera Flex	A	C16	Diploid Male	35.4	PRJNA107 5055
DiploidMale_14	Nextera Flex	A	C16	Diploid Male	34.4	PRJNA107 5055
DiploidMale_15	Nextera Flex	A	C1	Diploid Male	31.5	PRJNA107 5055
DiploidMale_16	Nextera Flex	A	C1	Diploid Male	32.2	PRJNA107 5055
LineC_DiploidFemale	Nextera Flex	C	?	Diploid Female	39.4	PRJNA107 5055
LineD_DiploidFemale_2	Nextera Flex	D	BG9	Diploid Female	35.2	PRJNA107 5055
LineD_DiploidFemale_1	Nextera Flex	D	?	Diploid Female	36.9	PRJNA107 5055
LineI_DiploidFemale	Nextera Flex	I	?	Diploid Female	36.1	PRJNA107 5055
LineM_DiploidFemale	Nextera Flex	M	?	Diploid Female	32.3	PRJNA107 5055
LineB_HaploidMale	Nextera Flex	B	?	Haploid Male	28.3	PRJNA107 5055
LineC_HaploidMale	Nextera	C	C9	Haploid	27.6	PRJNA107

	Flex			Male		5055
LineD_HaploidMale_1	Nextera Flex	D	BG7	Haploid Male	25.2	PRJNA1075055
LineD_HaploidMale_2	Nextera Flex	D	BG9	Haploid Male	34.9	PRJNA1075055
LineI_HaploidMale	Nextera Flex	I	BG2	Haploid Male	21.2	PRJNA1075055
LineM_HaploidMale_1	Nextera Flex	M	BG14	Haploid Male	70.7	PRJNA1075055
LineM_HaploidMale_2	Nextera Flex	M	BG14	Haploid Male	34.0	PRJNA1075055
LineB_DiploidFemale	Nextera Flex	B	STC6	Diploid Female	42.8	PRJNA1075055
ShortReadPedigree1_Daughter	Nextera Flex	B	STC6 – dsRed	Diploid Female	39.1	PRJNA947942
ShortReadPedigree1_Mother	Nextera Flex	B	STC6 – dsRed	Diploid Female	41.4	PRJNA947942
ShortReadPedigree2_Daughter	Nextera Flex	B	STC6 – dsRed	Diploid Female	42	PRJNA947942
ShortReadPedigree2_Mother	Nextera Flex	B	STC6 – dsRed	Diploid Female	42.7	PRJNA947942
LinkedReadPedigree2_Sister_1	TELL-Seq	B	STC6 – dsRed	Diploid Female	74.7	PRJNA947942
LinkedReadPedigree2_Sister_2	TELL-Seq	B	STC6 – dsRed	Diploid Female	196.7	PRJNA947942
LinkedReadPedigree1_Mother	TELL-Seq	B	STC6 – dsRed	Diploid Female	221.5	PRJNA947942
LinkedReadPedigree1_Daughter	TELL-Seq	B	STC6 – dsRed	Diploid Female	90	PRJNA947942
LinkedReadPedigree3_Daughter_1	TELL-Seq	B	STC6 – dsRed	Diploid Female	73.6	PRJNA947942
LinkedReadPedigree3_Mother	TELL-Seq	B	STC6 – dsRed	Diploid Female	50.1	PRJNA947942
LinkedReadPedigree3_Daughter_2	TELL-Seq	B	STC6 – dsRed	Diploid Female	100.8	PRJNA947942
LinkedReadPedigree4_Mother	TELL-Seq	B	STC6 – dsRed	Diploid Female	92.4	PRJNA947942
LinkedReadPedigree4_Daughter_1	TELL-Seq	B	STC6 – dsRed	Diploid Female	91.1	PRJNA947942
LinkedReadPedigree4_Daughter_2	TELL-Seq	B	STC6 – dsRed	Diploid Female	81	PRJNA947942
LinkedReadPedigree5_Sister_1	TELL-Seq	B	STC6 – dsRed	Diploid Female	75.5	PRJNA947942
LinkedReadPedigree5_Sister_2	TELL-Seq	B	STC6 – dsRed	Diploid Female	70	PRJNA947942
LinkedReadPedigree6_Sis	TELL-	B	STC6 –	Diploid	77.6	PRJNA947

ter_1	Seq		dsRed	Female		942
LinkedReadPedigree6_Sister_2	TELL-Seq	B	STC6 – dsRed	Diploid Female	69.3	PRJNA947942
LinkedReadPedigree7_Sister_1	TELL-Seq	B	STC6 - dsRed	Diploid Female	91.9	PRJNA947942
LinkedReadPedigree7_Sister_2	TELL-Seq	B	STC6 - dsRed	Diploid Female	82.3	PRJNA947942

Supplementary Table S2. Metadata for all DNA whole-genome shotgun sequencing libraries included in this study. Question marks indicate male samples for which the clonal line is known, but the stock colony of origin was not recorded.

L. fabarum CSD "region"	*L. fabarum* Gene ID	*O. biroi* homolog	*O. biroi* genomic locus
2	LF001188	XP_026823582.1	Chr2:8902488-8903163
2	LF001187	XP_011341081.1	Chr11:370661-370885
2	LF001186	XP_011337689.1	Chr9:12701112-12701246
2	LF001185	XP_011337690.1	Chr9:12701868-12702623
2	LF001185	XP_011337690.1	Chr9:12701868-12702623
2	LF001184	XP_019887284.1	Chr9:12784516-12784735
2	LF001183	XP_011337709.1	Chr9:12786107-12786502
2	LF001182	XP_011337688.2	Chr9:12691705-12691868
2	LF001182	XP_011337688.2	Chr9:12691705-12691868
2	LF001178	XP_011342643.1	Chr4:17418475-17418687
2	LF001175	XP_026825191.1	Chr4:17449644-17450001
2	LF001174	XP_011345175.1	Chr4:2098225-2098482
2	LF001173	XP_011351264.2	Chr4:17451378-17451779
2	LF001172	XP_011348371.2	Chr3:4120987-4121232
2	LF001171	XP_019886864.2	Chr4:13108487-13108728
2	LF001170	XP_011335781.1	Chr4:13104605-13104701
2	LF001169	XP_026824725.1	Chr4:13096987-13097135
2	LF001169	XP_026824725.1	Chr4:13096987-13097135
2	LF001168	XP_011334862.1	Chr1:5429156-5429315
2	LF001167	XP_011346080.1	Chr4:15317225-15317286
2	LF001166	XP_011346280.1	Chr4:15323952-15324443
2	LF001165	XP_011346265.1	Chr4:15325281-15325355
2	LF001164	XP_011333902.1	Chr9:13671279-13671529
2	LF001163	XP_011342394.2	Chr12:8945912-8946647
2	LF001162	XP_011347542.2	Chr5:13878906-13879057
2	LF001160	XP_011342397.1	Chr12:8898038-8898047
2	LF001159	XP_011342399.1	Chr12:8838643-8839361
2	LF001158	XP_026826742.1	Chr7:9069620-9070081
2	LF001157	XP_011336745.2	Chr7:9059981-9060079
2	LF001156	XP_011342557.1	Chr7:8937548-8938389
2	LF001155	XP_011333056.1	Chr2:8038775-8039241
2	LF001154	XP_011333053.1	Chr2:8074012-8074175
2	LF001153	XP_011333054.1	Chr2:8074146-8074195
2	LF001152	XP_011333051.1	Chr2:8149348-8149419
2	LF001151	XP_026827816.1	Chr8:9271044-9271606
2	LF001149	XP_026825880.1	Chr6:3921592-3922644
2	LF001148	XP_011338311.1	Chr7:9788316-9790754
2	LF001147	XP_011347933.1	Chr12:10190445-10190537
2	LF001146	XP_011335214.3	Chr11:13374668-13374686
2	LF001146	XP_011335214.3	Chr11:13374668-13374686
2	LF001146	XP_011335214.3	Chr11:13374668-13374686
2	LF001146	XP_011335214.3	Chr11:13374668-13374686
2	LF001146	XP_011335214.3	Chr11:13374668-13374686
2	LF001145	XP_026827302.1	Chr7:9709496-9709894
3	LF000382	XP_011346847.2	Chr1:10948255-10948488
3	LF000382	XP_011346847.2	Chr1:10948255-10948488
3	LF000383	XP_011346847.2	Chr1:10948255-10948488
3	LF000384	XP_011333915.1	Chr9:13583746-13584022
3	LF000386	XP_011338617.1	Chr2:15483180-15483410
3	LF000387	XP_011338616.2	Chr2:15484445-15484662
3	LF000390	XP_011338613.1	Chr2:15505270-15505403
3	LF000391	XP_026829310.1	Chr10:7183955-7184266
3	LF000392	XP_026826758.1	Chr7:5505525-5506433
3	LF000392	XP_026826758.1	Chr7:5505525-5506433
3	LF000393	XP_011341974.1	Chr7:5453558-5453828
3	LF000394	XP_019886900.1	Chr10:2038764-2039594
3	LF000395	XP_011349203.1	Chr4:4954193-4954244
3	LF000396	XP_011345154.1	Chr4:2157855-2158361

3	LF000397	XP_011352274.1	Chr7:203734-203802
3	LF000398	XP_026830084.1	Chr12:9934340-9935728
3	LF000399	XP_026830285.1	Chr13:8584249-8584824
3	LF000400	XP_011335143.1	Chr11:13391322-13391400
3	LF000401	XP_011336214.1	Chr11:11716747-11716790
3	LF000402	XP_011345301.1	Chr13:2249289-2249472
3	LF000403	XP_026830488.1	Chr13:2248176-2248407
3	LF000403	XP_026830488.1	Chr13:2248176-2248407
3	LF000403	XP_026830488.1	Chr13:2248176-2248407
3	LF000404	XP_011345919.1	Chr12:8096652-8097413
3	LF000405	XP_019889529.1	Chr13:9111829-9111940
3	LF000407	XP_011337549.1	Chr9:15451205-15451282
3	LF000408	XP_011351557.1	Chr5:4614020-4614156
3	LF000410	XP_011334973.2	Chr11:13498556-13498818
3	LF000411	XP_011337970.1	Chr7:12764736-12766373
3	LF002285	XP_011338781.1	Chr2:16117746-16118104
3	LF002286	XP_019887425.1	Chr2:16125070-16125404
3	LF002286	XP_019887425.1	Chr2:16125070-16125404
3	LF002287	XP_011338074.1	Chr7:12214725-12214961
3	LF002288	XP_011335962.1	Chr1:20844717-20844860
3	LF002289	XP_026823646.1	Chr2:16637198-16637380
3	LF002289	XP_026823646.1	Chr2:16637198-16637380
3	LF002289	XP_026823646.1	Chr2:16637198-16637380
3	LF002289	XP_026823646.1	Chr2:16637198-16637380
3	LF002290	XP_011336005.1	Chr1:20544946-20545045
3	LF002291	XP_026831057.1	Chr2:16812379-16813501
3	LF002292	XP_026831104.1	Chr2:16819356-16820705
3	LF002293	XP_026831104.1	Chr2:16819356-16820705
3	LF002296	XP_026828394.1	Chr9:14899993-14900040
3	LF002297	XP_019886768.2	Chr1:19222308-19222451
3	LF002297	XP_019886768.2	Chr1:19222308-19222451
3	LF002297	XP_019886768.2	Chr1:19222308-19222451
3	LF002298	XP_019889777.2	Chr7:5141809-5142057
3	LF002299	XP_019886808.1	Chr2:16884037-16884431
3	LF002300	XP_026825006.1	Chr4:2377212-2377305
3	LF002301	XP_011348694.1	Chr4:2367942-2368174
3	LF002302	XP_026825006.1	Chr4:2377212-2377305
3	LF002303	XP_011348699.1	Chr4:2386146-2386411
3	LF002304	XP_011350345.1	Chr4:8051873-8052036
3	LF002305	XP_011350347.1	Chr4:8052766-8052969
3	LF002306	XP_011337620.1	Chr9:15576650-15576680
3	LF002307	XP_011337625.1	Chr9:15576536-15576978
3	LF002308	XP_026824247.1	Chr3:10590193-10590478
3	LF002309	XP_019888231.1	Chr7:8924371-8925765
3	LF002310	XP_026824253.1	Chr3:10586163-10586300
3	LF002312	XP_026826755.1	Chr7:5509815-5510612
3	LF002313	XP_011329194.1	Chr11:847027-847118
3	LF002314	XP_011349559.2	Chr13:9386369-9386463
3	LF002315	XP_011333093.2	Chr3:9794768-9794937
3	LF002315	XP_011333093.2	Chr3:9794768-9794937
3	LF002318	XP_011333739.1	Chr3:15560244-15560328
3	LF002319	XP_011333739.1	Chr3:15560244-15560328
3	LF002321	XP_011344935.1	Chr10:9297465-9297664
3	LF002322	XP_011342759.1	Chr4:17388837-17389067
3	LF002323	XP_011333096.1	Chr3:9801594-9801632
3	LF002324	XP_026830338.1	Chr13:2378794-2378931
3	LF002325	XP_011333675.1	Chr11:9497118-9497256
3	LF002326	XP_011347567.1	Chr5:13957042-13957184
3	LF002327	XP_026830797.1	Chr14:7238938-7239145

3	LF002328	XP_011343665.1	Chr14:7321276-7321392
3	LF002328	XP_011343665.1	Chr14:7321276-7321392
3	LF002329	XP_011341939.1	Chr7:5246471-5246686
3	LF002330	XP_011347566.1	Chr5:13962033-13962259
3	LF002331	XP_026825510.1	Chr5:13964307-13964333
3	LF002332	XP_011344606.1	Chr1:22794512-22794678
3	LF002333	XP_011347655.1	Chr5:15247775-15247953
3	LF002334	XP_011341957.2	Chr7:5429102-5429536
3	LF002335	XP_011338783.2	Chr2:16112384-16112752
3	LF002336	XP_011335942.1	Chr1:21613665-21614096
3	LF002337	XP_026823940.1	Chr3:14533163-14533248
3	LF002338	XP_011333620.2	Chr3:14494369-14496705
3	LF002338	XP_011333620.2	Chr3:14494369-14496705
3	LF002340	XP_011333615.1	Chr3:14487295-14487394
3	LF002341	XP_011333610.1	Chr3:14469678-14470225
3	LF002342	XP_026831025.1	Chr1:1900527-1900742
3	LF002343	XP_011342722.1	Chr2:9443066-9443173
3	LF002344	XP_011336746.2	Chr7:9059981-9060079
3	LF002345	XP_026829935.1	Chr12:8103395-8103580
3	LF002345	XP_026829935.1	Chr12:8103395-8103580
3	LF002346	XP_011345921.1	Chr12:8107693-8108806
3	LF002347	XP_019888787.2	Chr12:8119273-8119816
3	LF002348	XP_011345923.2	Chr12:8124085-8124275
3	LF002348	XP_011345923.2	Chr12:8124085-8124275
3	LF002349	XP_026829855.1	Chr12:8157914-8158229
3	LF002349	XP_026829855.1	Chr12:8157914-8158229
3	LF002349	XP_026829855.1	Chr12:8157914-8158229
3	LF002350	XP_011345927.1	Chr12:8171391-8171651
3	LF002351	XP_011345928.2	Chr12:8172293-8172529
3	LF002353	XP_011347547.1	Chr5:13890987-13891126
3	LF002354	XP_026831106.1	Chr2:16898162-16900440
3	LF002355	XP_019890050.1	Chr10:1331247-1331374
3	LF002356	XP_011337159.1	Chr5:1700584-1700746
3	LF002357	XP_011352656.1	Chr11:2527061-2527567
3	LF002358	XP_011337758.2	Chr13:2376845-2376933
3	LF002359	XP_011342746.2	Chr4:17388313-17388616
3	LF002359	XP_011342746.2	Chr4:17388313-17388616
3	LF002360	XP_011343447.1	Chr4:17131792-17131946
3	LF002361	XP_011331147.2	Chr4:11068734-11068785
3	LF002362	XP_011345798.1	Chr4:15213481-15213677
3	LF002363	XP_011345817.1	Chr4:15215508-15215954
3	LF002364	XP_019889473.2	Chr4:15220312-15220911
3	LF002365	XP_011340809.1	Chr3:2701684-2702137
3	LF002366	XP_026824052.1	Chr3:2697959-2698230
3	LF002367	XP_026824725.1	Chr4:13096987-13097135
3	LF008641	XP_026824039.1	Chr3:9118311-9118866
4	LF005364	XP_011334504.1	Chr7:2098841-2099004
4	LF005365	XP_011342313.1	Chr12:9424093-9424190
4	LF005366	XP_011334041.1	Chr3:1349035-1349223
4	LF011889	XP_019886420.2	Chr3:10475465-10475573
5	LF003030	XP_026831077.1	Chr1:13554921-13555017
6	LF010888	XP_011344990.1	Chr10:8509474-8509767
6	LF009377	XP_011345605.2	Chr6:3471683-3472093
6	LF009377	XP_011345605.2	Chr6:3471683-3472093
6	LF009377	XP_011345605.2	Chr6:3471683-3472093
6	LF009379	XP_026828517.1	Chr9:8831304-8831370
6	LF009379	XP_026828517.1	Chr9:8831304-8831370
6	LF009379	XP_026828517.1	Chr9:8831304-8831370
6	LF009380	XP_011336797.1	Chr9:8837429-8837507

6	LF009381	XP_019887101.1	Chr9:8727410-8727519
6	LF009382	XP_026831406.1	Chr1:24511192-24511317
6	LF009382	XP_026831406.1	Chr1:24511192-24511317
6	LF009383	XP_026824186.1	Chr1:20167532-20167679
6	LF009384	XP_011342660.2	Chr7:8337474-8337620
6	LF009385	XP_011342660.2	Chr7:8337474-8337620
6	LF009386	XP_026828439.1	Chr9:8725506-8725902
6	LF009387	XP_026826104.1	Chr6:3608454-3608638
6	LF009389	XP_011345586.1	Chr6:3545693-3545764
6	LF009390	XP_011345587.1	Chr6:3542266-3542719
6	LF009391	XP_011336812.1	Chr9:9037569-9038229
6	LF009392	XP_026828409.1	Chr9:8861220-8865200
6	LF009393	XP_026829596.1	Chr11:13057235-13057309
6	LF009394	XP_011329564.1	Chr8:3674260-3674331
6	LF007431	XP_026828915.1	Chr1:10434755-10434933
6	LF007431	XP_026828915.1	Chr1:10434755-10434933
6	LF007432	XP_011331671.1	Chr9:14807644-14807839
6	LF007433	XP_011342213.1	Chr10:12611527-12611623
6	LF007434	XP_011342212.1	Chr10:12609888-12610342
6	LF007434	XP_011342212.1	Chr10:12609888-12610342
6	LF007436	XP_011337291.1	Chr2:9115215-9115236
6	LF007437	XP_026828146.1	Chr9:4587802-4588041
6	LF007438	XP_026829425.1	Chr11:6693825-6694146
6	LF007439	XP_011331274.1	Chr11:6694246-6694911
6	LF007440	XP_019886109.1	Chr11:6695786-6696124
6	LF007442	XP_011336194.1	Chr4:12380182-12380380
6	LF007443	XP_026830331.1	Chr13:9384275-9384535
6	LF007444	XP_011345852.1	Chr12:7367565-7367651
6	LF007445	XP_011335257.2	Chr3:7019336-7019555
6	LF007446	XP_026824186.1	Chr1:20167532-20167679
6	LF007446	XP_026824186.1	Chr1:20167532-20167679
6	LF007446	XP_026824186.1	Chr1:20167532-20167679
6	LF007446	XP_026824186.1	Chr1:20167532-20167679
6	LF007447	XP_026826373.1	Chr6:5911003-5911053
6	LF007448	XP_026824186.1	Chr1:20167532-20167679
6	LF007449	XP_011334216.1	Chr7:4769776-4770045
6	LF007450	XP_011349065.2	Chr12:27005-29771
6	LF007451	XP_011335419.1	Chr10:4847255-4847344
6	LF007452	XP_011335417.1	Chr10:4846815-4846859
6	LF007453	XP_026827807.1	Chr1:9089478-9089776
6	LF007455	XP_011335835.1	Chr10:2185332-2185676
6	LF007456	XP_011335834.1	Chr10:2178787-2178863
6	LF007457	XP_011329223.1	Chr11:762131-762397
6	LF007458	XP_011329224.1	Chr11:760857-760889
6	LF007459	XP_019886855.1	Chr4:13082208-13083182
6	LF007460	XP_026824731.1	Chr4:13078696-13078998
6	LF006413	XP_011335818.1	Chr10:2240022-2241763
6	LF006414	XP_011335862.1	Chr10:2135525-2135797
6	LF006415	XP_011335859.1	Chr10:2134826-2134936
6	LF006416	XP_011343295.1	Chr1:9058776-9058815
6	LF006417	XP_011343292.1	Chr1:9057219-9057283
6	LF006418	XP_011343292.1	Chr1:9057219-9057283
6	LF006419	XP_026829398.1	Chr11:2305329-2305581
6	LF006419	XP_026829398.1	Chr11:2305329-2305581
6	LF006420	XP_026829401.1	Chr11:2275861-2276530
6	LF006421	XP_011329296.1	Chr11:2271877-2272179
6	LF006422	XP_026829393.1	Chr11:7634951-7634983
6	LF006423	XP_026829706.1	Chr11:12285934-12286060
6	LF006424	XP_011344595.1	Chr1:22825831-22826081

6	LF006424	XP_011344595.1	Chr1:22825831-22826081
6	LF006425	XP_011352013.1	Chr8:6310826-6311037
6	LF006426	XP_011331632.1	Chr9:14579839-14579928
6	LF006427	XP_011352461.1	Chr7:15410331-15410681
6	LF006429	XP_011343549.1	Chr2:751822-752041
6	LF006430	XP_011349203.1	Chr4:4954193-4954244
6	LF006431	XP_011343549.1	Chr2:751822-752041
6	LF006432	XP_026827963.1	Chr9:14575261-14575420
6	LF006433	XP_011352574.1	Chr11:3070833-3070971
6	LF006433	XP_011352574.1	Chr11:3070833-3070971
6	LF006434	XP_011337828.2	Chr9:5812211-5812382
6	LF006435	XP_011349204.1	Chr4:4957779-4958419
6	LF006436	XP_011337810.1	Chr9:5882093-5882191
6	LF006437	XP_026823896.1	Chr3:805152-805214
6	LF006438	XP_011340786.1	Chr3:2598144-2598261
6	LF006439	XP_026823896.1	Chr3:805152-805214
6	LF006440	XP_011337822.1	Chr9:5860948-5861173
6	LF006440	XP_011337822.1	Chr9:5860948-5861173
6	LF006441	XP_019887296.1	Chr9:5860223-5860442
6	LF006442	XP_011331327.2	Chr11:5860356-5861681
6	LF006443	XP_026829555.1	Chr11:5855375-5855507
6	LF006444	XP_026824731.1	Chr4:13078696-13078998
6	LF004259	XP_011345653.1	Chr6:3026242-3028545
6	LF004259	XP_011345653.1	Chr6:3026242-3028545
6	LF004260	XP_011345657.1	Chr6:3023351-3023556
6	LF004261	XP_011333413.1	Chr3:11511683-11511793
6	LF004262	XP_011351420.1	Chr5:6432092-6432208
6	LF004264	XP_026826171.1	Chr6:5325828-5326190
6	LF004265	XP_011346014.1	Chr6:5325587-5325696
6	LF004266	XP_026826168.1	Chr6:5323418-5323504
6	LF004267	XP_011345652.1	Chr6:3045096-3045236
6	LF004269	XP_011345654.1	Chr6:3040496-3040563
6	LF004270	XP_011339354.1	Chr1:3930870-3931115
6	LF004271	XP_026826524.1	Chr6:3008743-3016188
6	LF004271	XP_026826524.1	Chr6:3008743-3016188
6	LF004271	XP_026826524.1	Chr6:3008743-3016188
6	LF004271	XP_026826524.1	Chr6:3008743-3016188
6	LF004272	XP_026828184.1	Chr9:13111407-13111541
6	LF004273	XP_026828185.1	Chr9:13086623-13086800
6	LF004274	XP_011333961.2	Chr9:13105804-13106034
6	LF004275	XP_011348142.2	Chr4:1434786-1435024
6	LF004276	XP_011340055.1	Chr13:7326557-7326826
6	LF004277	XP_026830287.1	Chr13:5035511-5035847
6	LF004279	XP_019887713.2	Chr13:4920080-4920324
6	LF004280	XP_011340067.2	Chr13:4897303-4901754
6	LF004280	XP_011340067.2	Chr13:4897303-4901754
6	LF004280	XP_011340067.2	Chr13:4897303-4901754
6	LF004281	XP_011340070.1	Chr13:4895933-4896176
6	LF004282	XP_011336869.1	Chr9:7349357-7349991
6	LF004283	XP_011339358.1	Chr1:3933442-3933591
6	LF004285	XP_011339356.1	Chr1:3937027-3937044
6	LF004286	XP_011339359.1	Chr1:3941264-3941427
6	LF004287	XP_011339359.1	Chr1:3941264-3941427
6	LF004288	XP_011344952.1	Chr10:9007970-9008172
6	LF004289	XP_011344954.1	Chr10:9003166-9003200
6	LF004290	XP_011344958.1	Chr10:8994570-8994686
6	LF004291	XP_011329385.1	Chr10:1633362-1633470
6	LF004292	XP_011329393.1	Chr10:1634963-1635204
6	LF004293	XP_011329393.1	Chr10:1634963-1635204

6	LF004294	XP_011329389.1	Chr10:1645838-1646248
6	LF004295	XP_026829098.1	Chr10:1647341-1647849
6	LF004296	XP_011329398.1	Chr10:1650088-1651341
6	LF004296	XP_011329398.1	Chr10:1650088-1651341
6	LF004297	XP_011337200.1	Chr7:7915393-7915548
6	LF004298	XP_026827348.1	Chr7:7916367-7916552
6	LF004299	XP_026827076.1	Chr1:7279248-7279325
6	LF004300	XP_011345863.1	Chr12:7388227-7388336
6	LF004301	XP_011345866.1	Chr12:7391418-7391629
6	LF004302	XP_011345466.1	Chr2:14533815-14534021
6	LF004303	XP_011343443.1	Chr7:10202220-10202623
6	LF004305	XP_011329969.1	Chr8:13384806-13384905
6	LF004305	XP_011329969.1	Chr8:13384806-13384905
6	LF004306	XP_026825681.1	Chr5:10051561-10051588
6	LF004307	XP_011346998.2	Chr8:14292863-14293716
6	LF004308	XP_011337232.1	Chr7:7682544-7682629
6	LF004309	XP_011331820.1	Chr9:450924-451223
6	LF004310	XP_011330274.1	Chr4:10793001-10793258
6	LF004311	XP_011345060.1	Chr10:7937737-7937849
6	LF004312	XP_026827788.1	Chr1:9083319-9083473
6	LF004312	XP_026827788.1	Chr1:9083319-9083473
6	LF004313	XP_011343304.1	Chr1:9082377-9082587
6	LF004314	XP_011330198.1	Chr4:10755815-10755993
6	LF004315	XP_011334459.1	Chr7:2283750-2284486
6	LF004316	XP_011339244.1	Chr1:3588029-3588192
6	LF004317	XP_026828329.1	Chr9:14054627-14054801
6	LF004317	XP_026828329.1	Chr9:14054627-14054801
6	LF004318	XP_011331508.2	Chr9:14061763-14061837
6	LF004319	XP_011331505.1	Chr9:14062698-14062811
6	LF004320	XP_011334857.1	Chr1:5454703-5454882
6	LF005974	XP_026828018.1	Chr1:5439229-5439530
6	LF005975	XP_011334860.2	Chr1:5434057-5434711
6	LF005976	XP_026827305.1	Chr7:8473920-8474183
6	LF005977	XP_011330739.1	Chr8:10462228-10462411
6	LF005978	XP_011330737.1	Chr8:10468312-10468396
6	LF005979	XP_011349122.1	Chr12:277094-277179
6	LF005980	XP_026824516.1	Chr3:12947714-12949000
6	LF005981	XP_011332350.1	Chr6:1628389-1628441
6	LF005982	XP_011351118.1	Chr14:2865850-2866011
6	LF005983	XP_011345061.1	Chr10:7928695-7928868
6	LF005984	XP_011345062.1	Chr10:7924312-7924663
6	LF005985	XP_011338878.1	Chr1:1486009-1486266
6	LF005986	XP_011345091.1	Chr10:7358649-7358818
6	LF005987	XP_011345085.1	Chr10:7382381-7382723
6	LF005989	XP_011345087.1	Chr10:7383338-7383519
6	LF005991	XP_019888652.1	Chr10:7645163-7645175
6	LF005992	XP_011345080.1	Chr10:7647698-7647862
6	LF005993	XP_011345081.1	Chr10:7649238-7649498
6	LF005994	XP_011335587.1	Chr10:3611073-3611430
6	LF005995	XP_011335585.1	Chr10:3617454-3617666
6	LF005996	XP_011335586.1	Chr10:3618392-3618466
6	LF005998	XP_011335579.1	Chr10:3670372-3670597
6	LF005999	XP_011335574.1	Chr10:3798053-3798466
6	LF006000	XP_011335569.1	Chr10:3800730-3800831
6	LF006002	XP_019888003.1	Chr6:2510924-2511178
6	LF006003	XP_011341632.2	Chr6:2491411-2491420
6	LF006004	XP_011341633.1	Chr6:2485102-2485164
6	LF006005	XP_026826688.1	Chr6:2483634-2483969
6	LF006006	XP_011334809.1	Chr1:5682988-5683224

6	LF006007	XP_011333725.1	Chr3:15156949-15157037
6	LF006008	XP_026827630.1	Chr1:5696215-5696343
6	LF006009	XP_011334805.1	Chr1:5696804-5696984
6	LF006011	XP_026827900.1	Chr8:3473774-3473863
6	LF006012	XP_011346986.1	Chr8:14327158-14327217
6	LF006013	XP_011346817.1	Chr1:11047741-11047827
6	LF006014	XP_011346814.2	Chr1:11042895-11042925
6	LF006015	XP_011343298.1	Chr1:9066719-9067238
6	LF006016	XP_011343299.1	Chr1:9074220-9074383
6	LF006018	XP_011330337.1	Chr11:12610357-12610821
6	LF006018	XP_011330337.1	Chr11:12610357-12610821
6	LF006019	XP_011330537.1	Chr4:10831178-10831370
6	LF006020	XP_011330547.1	Chr4:10830112-10830472
6	LF006599	XP_026824000.1	Chr3:15073019-15073218
6	LF006601	XP_011333707.2	Chr3:15069612-15069744
6	LF006602	XP_026824000.1	Chr3:15073019-15073218
6	LF006602	XP_026824000.1	Chr3:15073019-15073218
6	LF006602	XP_026824000.1	Chr3:15073019-15073218
6	LF006602	XP_026824000.1	Chr3:15073019-15073218
6	LF006602	XP_026824000.1	Chr3:15073019-15073218
6	LF006603	XP_011333707.2	Chr3:15069612-15069744
6	LF006603	XP_011333707.2	Chr3:15069612-15069744
6	LF006604	XP_011333707.2	Chr3:15069612-15069744
6	LF006604	XP_011333707.2	Chr3:15069612-15069744
6	LF006605	XP_026824000.1	Chr3:15073019-15073218
6	LF006606	XP_011331848.1	Chr9:587811-587881
6	LF006607	XP_011333716.1	Chr3:15073981-15074049
6	LF006608	XP_011329905.2	Chr8:14088136-14088231
6	LF006609	XP_011329905.2	Chr8:14088136-14088231
6	LF006611	XP_011334791.1	Chr1:5731015-5731140
6	LF006613	XP_011334796.1	Chr1:5726449-5726931
6	LF006614	XP_011332839.1	Chr6:14149691-14150234
6	LF006615	XP_011332839.1	Chr6:14149691-14150234
6	LF006616	XP_011332821.2	Chr6:14152496-14152628
6	LF006616	XP_011332821.2	Chr6:14152496-14152628
6	LF006616	XP_011332821.2	Chr6:14152496-14152628
6	LF006617	XP_011332825.1	Chr6:14184424-14184557
6	LF006618	XP_011332827.2	Chr6:14191487-14191902
6	LF006619	XP_011332863.2	Chr6:14063075-14064223
6	LF006620	XP_011332866.3	Chr6:14083471-14083668
6	LF006621	XP_011332857.1	Chr6:14084068-14084244

<u>Supplementary Table S3.</u> *O. biroi* orthologs of genes located in the putative *L. fabarum* CSD "regions".

V. emeryi **CSD QTL**	***V. emeryi*** **GeneID**	**OrthoFinder** ***O. biroi*** **"Ortholog"**	***O. biroi*** **genomic locus**
2	LOC105559778	LOC105285603	Chr4:1637384-1672949
2	LOC105559779	LOC105285604	Chr4:1673043-1674418
2	LOC105559781	LOC105285607	Chr4:1674523-1676002
2	LOC105559780	LOC105285605	Chr4:1675932-1678296
2	LOC105559782	LOC105283856, LOC113561733	Chr4:1727256-1729679; Chr4:22762-24749
2	LOC105559784	LOC105283852	Chr4:1729888-1737844
2	LOC105559783	LOC105283851	Chr4:1730326-1732211
2	LOC105559785	LOC105283850	Chr4:1738076-1749057
2	LOC105559786	LOC105282554	Chr4:7069030-7075460
2	LOC105559787	LOC105282557	Chr4:7075106-7080018
2	LOC105559788	LOC105286165	Chr4:4951574-4955391
2	LOC105559790	LOC105282568	Chr4:7141680-7152824
2	LOC105559789	LOC105282568	Chr4:7141680-7152824
2	LOC105559791	LOC105283915	Chr4:5377245-5381427
2	LOC105559793	LOC105283797	Chr4:2173362-2237652
1	LOC105558826	LOC105287916	Chr2:985469-993787
1	LOC105558815	LOC105283998	Chr2:14515051-14517463
1	LOC105558816	LOC105283997, LOC113563596	Chr2:14517342-14519998; Chr2:19219509-19221771
1	LOC105558817	LOC105283996, LOC113563597	Chr2:14519947-14522109; Chr2:19217723-19219230
1	LOC105558818	LOC105283999	Chr2:14522233-14522882
1	LOC105558819	LOC105278142	Chr2:16951611-16953740
1	LOC105558820	LOC105278143	Chr2:16947572-16951531
1	LOC105558821	LOC105278141	Chr2:16939029-16945082
1	LOC105558822	LOC105287131	Chr2:12765464-12768244
1	LOC105558823	LOC105278140	Chr2:16929798-16936497

<u>Supplementary Table S4.</u> *O. biroi* orthologs of genes located in the two *V. emeryi* CSD QTLs.

	Gene_type	RU	RefSeq	GenBank
Genes	guide_RNA	0	1	0
	lncRNA	1767	1430	0
	miRNA	239	0	0
	misc_RNA	0	44	0
	piRNA	23596	0	0
	protein_coding	12839	11927	13022
	pseudogene	0	373	77
	rRNA	108	109	0
	snoRNA	15	16	0
	snRNA	22	26	0
	tRNA	194	202	0
Total Genes		**38780**	**14128**	**13099**
Transcripts	lncRNA	3795	1751	0
	miRNA	239	0	0
	mRNA	63620	24215	13099
	piRNA	23596	0	0
	rRNA	108	109	0
	snoRNA	15	16	0
	snRNA	22	26	0
	tRNA	194	202	0
	guide_RNA	0	1	0
	transcript	0	818	0
Total transcripts		**91589**	**27135**	**13099**
Total multi-exon transcripts		**65882**	**25945**	**11923**

<u>Supplementary Table S5.</u> Improvements over previous genome annotations, with the new (RU) annotation featuring genes and transcripts not found in previous annotation versions.

Gene Name	Gene ID	Log 2-Fold Change	p-value	p-adjusted
unchar. protein	LOC105285599	0.11	0.84	0.92
unchar. protein	LOC105285600	0.11	0.73	0.85
paraneoplastic antigen Ma6E	LOC105285601	0.21	0.67	0.82
ataxin-7	LOC105285602	-0.33	0.063	0.17
Kank	LOC105285603	0.05	0.82	0.91
Bet1	LOC105285604	0.08	0.54	0.72
prob. ribosome prod. factor	LOC105285607	0.34	0.12	0.27
THUMP	LOC105285605	0.19	0.13	0.28
unchar. lncRNA	LOC105285608	0.12	0.87	0.93
Creld	LOC113561733	0.56	0.77	0.88
alphaCOP	LOC105283852	-0.17	0.21	0.40
unchar. protein	LOC105283851	9.95	NA	NA
Host cell factor	LOC105283850	-0.52	2.1E-04	1.8E-03
l(2)37Cg	LOC105283849	0.86	5.7E-04	4.2E-03
ribosomal protein S2	LOC105283847	-0.06	0.50	0.69
guanine deaminase	LOC105283846	0.03	0.88	0.94
unchar. protein	LOC105283844	1.77	6.1E-08	1.0E-07

Supplementary Table S6. Differential gene expression near the CSD index peak. The log 2-fold change in expression levels between male and female samples, p-values (Wald test), and adjusted p-values (Benjamini-Hochberg adjustment) for 17 genes in the vicinity of the CSD index peak. The p-values are “NA” for LOC105283850 because one of the samples is an extreme outlier detected by Cook’s distance.

Gene Name	Gene ID	Exon	Log 2-Fold Change	p-value	p-adjusted
unchar. protein	LOC105285599	E001	-0.24	0.37	0.84
		E002	0.09	0.6	0.91
		E003	0.21	0.63	0.92
unchar. protein	LOC105285600	E001	-0.59	0.32	0.82
		E002	-0.23	0.78	0.96
		E003	0.51	0.15	0.72
		E004	0.03	0.87	0.98
		E005	0.06	0.73	0.95
		E006	-0.54	0.31	0.82
		E007	-0.48	0.35	0.83
		E008	0.11	0.91	0.99
		E009	-0.66	0.65	0.92
		E010	0.08	0.92	0.99
		E011	0.11	0.91	0.99
paraneoplastic antigen Ma6E	LOC105285601	E001	-0.28	0.06	0.51
		E002	0.18	0.06	0.5
ataxin-7	LOC105285602	E001	2.03	0.1	0.63
		E002	0.12	0.98	1
		E003	0.06	0.75	0.95
		E004	0.05	0.69	0.94
		E005	-0.04	0.48	0.87
		E006	-0.48	0.01	0.17
		E007	0.12	0.55	0.89
Kank	LOC105285603	E001	-0.93	0.24	0.79
		E002	-0.65	0.05	0.47
		E003	-0.14	0.72	0.94
		E004	-1.04	0.44	0.86
		E005	-1.6	0.29	0.81
		E006	-0.55	0.08	0.57
		E007	-0.67	0.04	0.4
		E008	-0.63	0.08	0.58
		E009	0.01	0.88	0.98
		E010	-0.15	0.34	0.83
		E011	-0.28	0.2	0.76
		E012	-0.27	0.2	0.76
		E013	-0.27	0.18	0.74
		E014	0.07	0.3	0.81

		E015	0.11	0.07	0.54
		E016	0.18	0.26	0.8
		E017	NA	NA	NA
Bet1	LOC105285604	E001	2.7	0.05	0.47
		E002	0	0.99	1
		E003	-0.04	0.67	0.93
		E004	-0.01	0.82	0.97
		E005	0.05	0.94	1
		E006	9.44	0.23	0.78
prob. ribosome prod. factor	LOC105285607	E001	0.24	0.77	0.96
		E002	-0.12	0.77	0.96
		E003	0.05	0.7	0.94
		E004	0	1	1
		E005	0.3	0.55	0.89
THUMP	LOC105285605	E001	1.26	0.04	0.42
		E002	-0.12	0.49	0.87
		E003	-0.09	0.55	0.89
		E004	0.03	0.79	0.96
		E005	0.13	0.19	0.75
		E006	-0.03	0.96	1
		E007	-0.09	0.76	0.95
		E008	-0.09	0.68	0.93
		E009	3.9	0.000257	0.00951
unchar. lncRNA	LOC105285608	E001	-1.46	0.08	0.56
		E002	0.28	0.06	0.5
Creld	LOC113561733	E001	-0.06	0.62	0.92
		E002	-0.01	0.86	0.98
		E003	0.09	0.32	0.82
		E004	-0.15	0.6	0.91
alphaCOP	LOC105283852	E001	-0.29	0.3	0.81
		E002	-0.23	0.33	0.83
		E003	-0.21	0.25	0.79
		E004	-0.15	0.16	0.73
		E005	0.14	0.19	0.75
		E006	0.32	0.3	0.81
		E007	0.74	0.14	0.69
unchar. protein	LOC105283851	E001	-0.07	0.76	0.96
		E002	0.14	0.76	0.96
Host cell factor	LOC105283850	E001	-0.34	0.28	0.8

		E002	-0.33	0.24	0.79
		E003	-0.25	0.38	0.84
		E004	-0.34	0.36	0.84
		E005	-0.19	0.5	0.88
		E006	-0.19	0.39	0.85
		E007	0.75	0.12	0.67
		E008	-0.23	0.33	0.82
		E009	0.3	0.39	0.85
		E010	-0.05	0.93	1
		E011	0.44	0.34	0.83
		E012	0	0.59	0.91
		E013	0	0.35	0.83
		E014	0.45	0.12	0.67
		E015	0.04	0.2	0.77
		E016	-0.35	0.56	0.9
		E017	0.14	0.07	0.55
		E018	0.11	0.48	0.87
		E019	0.35	0.38	0.84
l(2)37Cg	LOC105283849	E001	0.04	0.54	0.89
		E002	-0.03	0.6	0.91
		E003	0	0.94	1
ribosomal protein S2	LOC105283847	E001	0.1	0.54	0.89
		E002	-0.03	0.45	0.86
		E003	-0.06	0.63	0.92
		E004	3.68	0.00233	0.06
guanine deaminase	LOC105283846	E001	0.29	0.31	0.82
		E002	0.15	0.52	0.89
		E003	0.15	0.37	0.84
		E004	0.12	0.55	0.89
		E005	0.1	0.58	0.9
		E006	-0.06	0.62	0.92
		E007	-0.09	0.41	0.85
		E008	-0.17	0.18	0.75
		E009	-0.26	0.07	0.54
unchar. protein	LOC105283844	E001	-0.52	0.000763	0.02
		E002	-0.54	0.00000519	0.000311
		E003	-0.56	0.000378	0.01
		E004	-0.91	0.00693	0.13
		E005	-0.85	0.00588	0.11

		E006	-0.92	0.000107	0.00449
		E007	-0.91	0.00434	0.09
		E008	-0.8	0.00161	0.04
		E009	-0.53	0.00458	0.1
		E010	-0.54	0.0000288	0.00143
		E011	-0.46	0.000237	0.00887
		E012	0.57	7.31E-07	0.0000543
		E013	0.79	0.00000443	0.00027
		E014	1.11	1.32E-07	0.0000115
		E015	1.07	3.96E-07	0.000031
		E016	0.86	0.000705	0.0219
		E017	0.8	0.000814	0.0245
		E018	1.29	0.000153	0.00613
		E019	1.29	0.000454	0.0152
		E020	1.13	0.000177	0.00694
		E021	0.99	0.00261	0.06
		E022	-0.89	0.55	0.89
		E023	1.32	0.000135	0.00547
		E024	1.11	0.07	0.55
		E025	-0.57	0.44	0.86
		E026	-0.24	0.32	0.82
		E027	0.09	0.00973	0.165
		E028	0.21	0.0003	0.0107
		E029	-0.59	2.27E-08	0.00000229
		E030	-0.23	0.00000842	0.000479
		E031	0.51	0.00158	0.0418

<u>Supplementary Table S7.</u> Differential gene expression near the CSD index peak.

The log 2-fold change in expression levels between male and female samples, p-values (likelihood ratio test), and adjusted p-values (Benjamini-Hochberg adjustment) for each exon of 17 genes in the vicinity of the CSD index peak. The p-values are “NA” for one of the exons of LOC105285603 because there is negligible expression of that exon (basemean=0).

Analysis	**Program**	**Version**	**Analysis Available At:**
Genetic Mapping	Trimmomatic	v0.36	https://github.com/kipdlacy/SexDeterminationMapping_Obiroi
Genetic Mapping	bwa	v0.7.12-r1039	https://github.com/kipdlacy/SexDeterminationMapping_Obiroi
Genetic Mapping	GATK	v4.2	https://github.com/kipdlacy/SexDeterminationMapping_Obiroi
Genetic Mapping	vcftools	v0.1.17	https://github.com/kipdlacy/SexDeterminationMapping_Obiroi
Genetic Mapping	pyvcf	v0.6.8	https://github.com/kipdlacy/SexDeterminationMapping_Obiroi
Genetic Mapping	pybedtools	v0.9.0	https://github.com/kipdlacy/SexDeterminationMapping_Obiroi
Genetic Mapping	pandas	v1.1.5	https://github.com/kipdlacy/SexDeterminationMapping_Obiroi
Genetic Mapping	numpy	v.1.19.5	https://github.com/kipdlacy/SexDeterminationMapping_Obiroi
Genetic Mapping	bcftools	v0.1.19-96b5f2294a	https://github.com/kipdlacy/SexDeterminationMapping_Obiroi
Genetic Mapping	samtools	v1.8	https://github.com/kipdlacy/SexDeterminationMapping_Obiroi
Genetic Mapping	OrthoFinder	v2.5.5	https://github.com/kipdlacy/SexDeterminationMapping_Obiroi
Genetic Mapping	scikit-allel	v1.3.5	https://github.com/kipdlacy/SexDeterminationMapping_Obiroi
Genetic Mapping	blast	v2.5.0	https://github.com/kipdlacy/SexDeterminationMapping_Obiroi
Genetic Mapping	mummer	v4.0.0rc1	https://github.com/kipdlacy/SexDeterminationMapping_Obiroi
Genetic Mapping	python	v3.6.15	https://github.com/kipdlacy/SexDeterminationMapping_Obiroi
Genetic Mapping	R	v4.3.2	https://github.com/kipdlacy/SexDeterminationMapping_Obiroi
Genetic Mapping	karyoploteR	1.28.0	https://github.com/kipdlacy/SexDeterminationMapping_Obiroi
Genetic Mapping	ggplot2	v3.5.0	https://github.com/kipdlacy/SexDeterminationMapping_Obiroi
Genetic Mapping	ggbeeswarm	v0.7.2	https://github.com/kipdlacy/SexDeterminationMapping_Obiroi
Genetic Mapping	matplotlib	v3.3.4	https://github.com/kipdlacy/SexDeterminationMapping_Obiroi
Genome Annotation	pbmm2	v1.12.0	https://github.com/RockefellerUniversity/Obiroi_GeneModels_2025
Genome	stringtie2	v2.2.1	https://github.com/RockefellerUniversity/Obiroi_Ge

Annotation			neModels_2025
Genome Annotation	samtools	v1.6	https://github.com/RockefellerUniversity/Obiroi_GeneModels_2025
Genome Annotation	hisat2	v2.2.1	https://github.com/RockefellerUniversity/Obiroi_GeneModels_2025
Genome Annotation	transdecoder	v5.7.1	https://github.com/RockefellerUniversity/Obiroi_GeneModels_2025
Genome Annotation	squanti3	v3.5.1	https://github.com/RockefellerUniversity/Obiroi_GeneModels_2025
Genome Annotation	Repeatmasker	v4.1.7-p1	https://github.com/RockefellerUniversity/Obiroi_GeneModels_2025
Genome Annotation	FEELnc	v0.2.1	https://github.com/RockefellerUniversity/Obiroi_GeneModels_2025
Genome Annotation	BUSCO	v5.4.7	https://github.com/RockefellerUniversity/Obiroi_GeneModels_2025
Genome Annotation	CLIPflexR	v0.1.20	https://github.com/RockefellerUniversity/Obiroi_GeneModels_2025
Genome Annotation	miRDeep2	v0.1.3	https://github.com/RockefellerUniversity/Obiroi_GeneModels_2025
Genome Annotation	Pilfer	v1.0	https://github.com/RockefellerUniversity/Obiroi_GeneModels_2025
Genome Annotation	R	v4.0	https://github.com/RockefellerUniversity/Obiroi_GeneModels_2025
Genome Annotation	Bioconductor	v3.19	https://github.com/RockefellerUniversity/Obiroi_GeneModels_2025
Differential Gene Expression	multiqc	v1.9	https://github.com/jina-leemon/CSD_biroi_male-female-RNA-seq
Differential Gene Expression	RSEM	v1.2.28	https://github.com/jina-leemon/CSD_biroi_male-female-RNA-seq
Differential Gene Expression	STAR	v2.5.0a	https://github.com/jina-leemon/CSD_biroi_male-female-RNA-seq
Differential Gene Expression	R	v4.3.3	https://github.com/jina-leemon/CSD_biroi_male-female-RNA-seq
Differential Gene Expression	GenomicFeatures	v1.52.2	https://github.com/jina-leemon/CSD_biroi_male-female-RNA-seq
Differential Gene Expression	tximport	v1.28.0	https://github.com/jina-leemon/CSD_biroi_male-female-RNA-seq
Differential Gene Expression	DESeq2	v1.40.2	https://github.com/jina-leemon/CSD_biroi_male-female-RNA-seq

Differential Gene Expression	EnhancedVolcano	v1.18.0	https://github.com/jina-leemon/CSD_biroi_male-female-RNA-seq
Differential Gene Expression	pheatmap	v1.0.12	https://github.com/jina-leemon/CSD_biroi_male-female-RNA-seq
Differential Gene Expression	dplyr	v1.1.4	https://github.com/jina-leemon/CSD_biroi_male-female-RNA-seq
Differential Gene Expression	reshape2	v1.4.4	https://github.com/jina-leemon/CSD_biroi_male-female-RNA-seq
Differential Gene Expression	tidyr	v1.3.1	https://github.com/jina-leemon/CSD_biroi_male-female-RNA-seq
Differential Gene Expression	tibble	v3.2.1	https://github.com/jina-leemon/CSD_biroi_male-female-RNA-seq
Differential Gene Expression	ggbreak	v.0.1.2	https://github.com/jina-leemon/CSD_biroi_male-female-RNA-seq
Differential Exon Usage	HTSeq	v2.0.5	https://github.com/jina-leemon/CSD_biroi_male-female-RNA-seq
Differential Exon Usage	DEXSeq	v.1.46.0	https://github.com/jina-leemon/CSD_biroi_male-female-RNA-seq
Differential Exon Usage	stageR	v.1.22.0	https://github.com/jina-leemon/CSD_biroi_male-female-RNA-seq

Supplementary Table S8. Versions of software used for genomics analyses.