

# The genomics and evolution of inter-sexual mimicry and female-limited polymorphisms in damselflies

Received: 14 March 2023

Accepted: 4 October 2023

Published online: 6 November 2023

 Check for updates

Beatriz Willink<sup>1,2</sup>✉, Kalle Tunström<sup>1</sup>, Sofie Nilén<sup>3</sup>, Rayan Chikhi<sup>4</sup>,  
Téo Lemane<sup>5</sup>, Michihiko Takahashi<sup>6,8</sup>, Yuma Takahashi<sup>7</sup>, Erik I. Svensson<sup>1,3</sup>  
& Christopher West Wheat<sup>1</sup>

Sex-limited morphs can provide profound insights into the evolution and genomic architecture of complex phenotypes. Inter-sexual mimicry is one particular type of sex-limited polymorphism in which a novel morph resembles the opposite sex. While inter-sexual mimics are known in both sexes and a diverse range of animals, their evolutionary origin is poorly understood. Here, we investigated the genomic basis of female-limited morphs and male mimicry in the common bluetail damselfly. Differential gene expression between morphs has been documented in damselflies, but no causal locus has been previously identified. We found that male mimicry originated in an ancestrally sexually dimorphic lineage in association with multiple structural changes, probably driven by transposable element activity. These changes resulted in ~900 kb of novel genomic content that is partly shared by male mimics in a close relative, indicating that male mimicry is a trans-species polymorphism. More recently, a third morph originated following the translocation of part of the male-mimicry sequence into a genomic position ~3.5 mb apart. We provide evidence of balancing selection maintaining male mimicry, in line with previous field population studies. Our results underscore how structural variants affecting a handful of potentially regulatory genes and morph-specific genes can give rise to novel and complex phenotypic polymorphisms.

Sexual dimorphism is one of the most fascinating forms of intra-specific phenotypic variation in animals. Sexes often differ in size and colour, as well as the presence of elaborated ornaments and weaponry. Theoretical and empirical studies over many decades have developed a detailed framework of sexual selection and sexual conflict, explaining why these differences arise and how they become encoded in sex differentiation

systems<sup>1–3</sup>. However, a growing number of examples of inter-sexual mimicry<sup>4–7</sup> suggest that sexual dimorphism can be evolutionarily fragile and quite dynamic. Inter-sexual mimicry has evolved in several lineages, when individuals of one sex gain a fitness advantage, usually frequency- or density-dependent, due to their resemblance to the opposite sex. For example, males who mimic females, as seen in the ruff

<sup>1</sup>Department of Zoology, Stockholm University, Stockholm, Sweden. <sup>2</sup>Department of Biological Sciences, National University of Singapore, Singapore, Singapore. <sup>3</sup>Department of Biology, Lund University, Lund, Sweden. <sup>4</sup>Sequence Bioinformatics, Institut Pasteur, Université Paris Cité, Paris, France. <sup>5</sup>University of Rennes, Inria, CNRS, IRISA, Rennes, France. <sup>6</sup>Graduate School of Life Sciences, Tohoku University, Sendai, Japan.

<sup>7</sup>Graduate School of Science, Chiba University, Chiba, Japan. <sup>8</sup>Present address: Graduate School of Agriculture, Kyoto University, Kyoto, Japan.

✉e-mail: [beatriz.willink@zoologi.su.se](mailto:beatriz.willink@zoologi.su.se)

(*Calidris pugnax*) and the Melanzona guppy (*Poecilia parae*), forgo courtship and 'sneak' copulations from dominant males<sup>4,5</sup>, while females who mimic males, in damselflies and hummingbirds, avoid excessive male-mating harassment<sup>6,8</sup>. Inter-sexual mimicry thus requires the evolution of a novel sex-mimicking morph in a sexually dimorphic ancestor. The occurrence of inter-sexual mimicry may be an intermediate step in the evolution of sexual monomorphism, it may be an ephemeral state or it may be maintained as a stable polymorphism. In any case, sexual mimics harbour genetic changes that attenuate or prevent the development of sex-specific phenotypes, and can therefore provide insights into the essential building blocks of sexual dimorphism<sup>9</sup>.

Considerable research effort has been devoted to uncover the genetic basis of discrete phenotypic polymorphisms, such as those associated with alternative reproductive or life-history strategies<sup>10–14</sup>. Together, these studies highlight a vast diversity of mechanisms used by evolution to package complex phenotypic differences into a single locus that is protected from the eroding effects of recombination. At one extreme, phenotypic morphs may evolve via massive insertions, deletions or inversions that lock together dozens to hundreds of genes into supergenes<sup>15–17</sup>. At the other end, much smaller structural variants (SVs), confined to a few thousand base pairs, can modulate the expression of one or a few regulators of pleiotropic networks, resulting in markedly different morphs<sup>11,12,18</sup>. We are clearly only starting to get a glimpse of the major themes among these genetic mechanisms. For example, it is not known whether genomic architecture determines the type and breadth of co-varying traits or the likelihood of polymorphisms evolving in specific lineages<sup>19</sup>.

A few of these studies have focused on sex-limited polymorphisms, where one of the morphs shares the overall appearance, such as the colour pattern, of the opposite sex<sup>10,14,20</sup>. Such sex-limited morphs may illustrate novel origins of sexual dimorphism, driven by either sexual selection in males<sup>14</sup> or natural selection in females<sup>18,21</sup>. Alternatively, sex-limited polymorphisms may arise with the evolution of inter-sexual mimicry. Crucially, empirical support for the evolution of inter-sexual mimicry demands both a macroevolutionary context for the polymorphism, showing that sexually dimorphism is ancestral, and a documented advantage of sexual mimics in at least some social contexts. There is therefore a need to integrate genomic, microevolutionary and phylogenetic evidence into our understanding of the evolutionary dynamics of sexual dimorphism and inter-sexual mimicry. This integrative approach has been overall rare, and applied mostly to the study of alternative male reproductive strategies<sup>18,22</sup>. Yet, female mimicry of males may be more common than historically appreciated<sup>23</sup>, and the genetic basis of such mimicry remains largely unexplored<sup>24–26</sup>.

The common bluetail damselfly *Ischnura elegans* (Odonata) has three female-limited morphs (namely O, A and I) that differ in colouration, whereas males are always monomorphic<sup>27</sup>. O females display the colour pattern and developmental colour changes inferred as ancestral in a comparative analysis of the genus *Ischnura*<sup>28</sup> (Fig. 1). Male-like (A) females are considered male mimics, who experience a frequency-dependent advantage of reduced male mating and pre-mating harassment due to their resemblance to males<sup>6</sup>. Finally, the I morph shares its stripe pattern and immature colouration with the A morph<sup>27</sup> (Fig. 1), but develops a yellow-brown background colouration with age, eventually resembling the O morph upon sexual maturation<sup>29</sup>. I females are only known in *I. elegans* and a few close relatives<sup>28</sup> (Fig. 1), and their evolutionary relationship to A and O females remains unresolved. The behaviour, ecology and population biology of *I. elegans* have been intensely investigated for over two decades, making it one of the best understood female-limited polymorphisms, in terms of how morphs differ in fitness-related traits and how alternative morphs are maintained sympatrically over long periods<sup>30–33</sup>. Nonetheless, the molecular basis of this polymorphism remains unknown.

To advance our understanding of the evolution of complex phenotypes, such as sexual dimorphism and sex-specific morphs,

we identify the genomic region responsible for the female-limited colour polymorphism in *I. elegans*. Using a combination of reference-based and reference-free genome-wide association studies (GWAS), upon morph-specific genome assemblies, we revealed two novel regions adding up to ~900 kb that are associated with the evolutionary origin of the male-mimicking A morph. These SVs, probably generated and expanded by transposable element (TE) activity, are partly shared by male-mimicking females of the tropical bluetail damselfly (*Ischnura senegalensis*), indicating that male mimicry is a trans-species polymorphism. We also show that the novel I morph evolved via an ectopic recombination event, where part of the A-unique genomic content was translocated into an O genomic background. Finally, we examined the evolutionary dynamics of the colour morph locus and explored expression patterns of genes located in this region. Together, our results indicate that structural variation affecting a handful of genes and maintained by balancing selection provides the raw material for the evolution of a male-mimicking phenotype in pond damselflies.

## Results

### Male mimicry is encoded by a locus with a signature of balancing selection

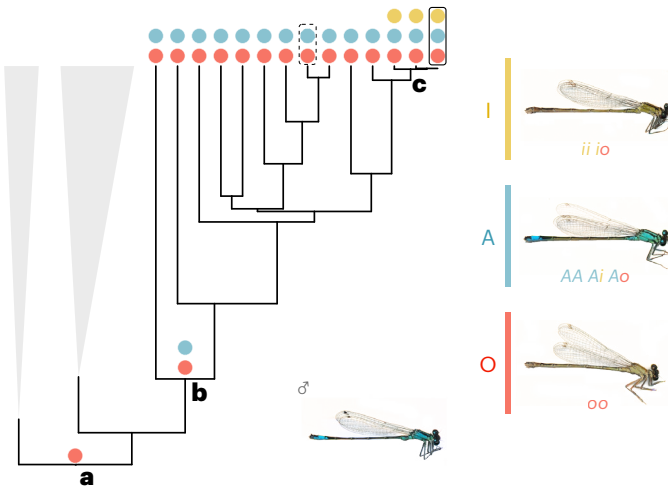
We started by conducting three reference-based GWAS, comparing all morphs against each other in a pairwise fashion (Extended Data Fig. 1). We used an A morph genome assembly (Supplementary Text 1) as mapping reference because SV analyses revealed that A females harbour genomic content that is absent in the other two morphs (see 'Female morphs differ in genomic content'). The draft assembly was scaffolded against the Darwin Tree of Life (DTOL) reference genome to place the contigs in a chromosome-level framework<sup>34</sup>. The DTOL reference genome contains the O allele (see Supplementary Text 2) and is assembled with chromosome resolution, except for chromosome 13, which is fragmented and consists of one main and several unlocalized scaffolds.

All pairwise GWAS between morphs pointed to one and the same unlocalized scaffold of chromosome 13 as the causal morph locus (Fig. 2a). Closer examination of this scaffold revealed two windows of elevated divergence between morphs (Fig. 2b). First, a narrow region near the start of the scaffold (~50 kb–0.2 mb) captures highly significant single nucleotide polymorphisms (SNPs) in both A versus O and I versus O comparisons (Fig. 2b). Thereafter, and up to ~1.5 mb, an abundance of SNPs differentiates A females from both O and I females, especially between ~0.6 and ~1.0 mb (Fig. 2b). These results are mirrored by genetic differentiation ( $F_{ST}$ ) values across both regions (Fig. 2c).

Next, we investigated whether the morph locus carries a signature of balancing selection, as suggested by previous field studies of morph-frequency dynamics<sup>31</sup>. The larger genomic window that uniquely distinguishes A females from both I and O females displays a signature of balancing selection, indicated by highly positive values of Tajima's  $D$ , exceeding the 95th percentile of genome-wide estimates (Fig. 2d). Conversely, values of both Tajima's  $D$  and nucleotide diversity ( $\pi$ ) in the narrower window that differentiates O females from both A and I females (~50 kb–0.2 mb) fall within the 95th percentile of genome-wide estimates (Fig. 2d–e).

### Female morphs differ in genomic content

Previous studies have found that complex phenotypic polymorphisms are often underpinned by SVs, arising from genomic rearrangements such as insertions, deletions and inversions<sup>10,13,15,20</sup>. As these variants can be difficult to detect in a reference-based analysis, we employed a  $k$ -mer based GWAS approach<sup>35</sup> (Extended Data Fig. 1), which enables reference-free identification of genomic divergence between morphs. Significant  $k$ -mers in these analyses could represent regions that are present in one morph and absent in the other (that is, insertions or deletions), or regions that are highly divergent in their sequence (as in a traditional GWAS).



**Fig. 1 | The evolution of female-limited colour polymorphisms in *Ischnura* damselflies.** **a**, A previous phylogenetic study and ancestral state reconstruction<sup>25</sup> proposed that the genus *Ischnura* had a sexually dimorphic ancestor, with O-like females (red circle). The O morph is markedly different from males, having a bronze-brown thorax and faint stripes, instead of the black thoracic stripes on a bright blue background of males. **b**, Male mimicry (A females, blue circle) has evolved more than once, for instance, in an ancestor of the (expanded) clade that includes the common bluetail (*I. elegans*, outlined with solid line) and the tropical bluetail (*I. senegalensis*, outlined with dashed line). **c**, *I. elegans* is a trimorphic species, due to the recent evolution of a third female morph, I (yellow circle). In *I. elegans*, morph inheritance follows a dominance hierarchy, where the most dominant allele produces the A morph and two copies of the most recessive allele are required for the development of O females. In contrast, the O allele is dominant in *I. senegalensis*<sup>106</sup>. Terminal nodes in the phylogeny represent different species. Grey triangles represent other clades of *Ischnura*, which are collapsed for clarity. Damselfly images adapted from ref. 25 under a Creative Commons licence [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/).

First, we investigated the divergence associated with the male-mimicking A morph. Pairwise analyses revealed 568,039 and 508,031 *k*-mers (length = 31 bp) significantly associated with the A versus O and A versus I comparisons, respectively. To determine whether the associated *k*-mers represent differences in genomic content or sequence between the morphs, we mapped these *k*-mers to morph-specific reference genomes. If the associated *k*-mers are owing to novel sequences found in one morph but not the other, we would expect a vast majority of the significant *k*-mers to be found in only one of the two morphs in a pairwise comparison. If the significant *k*-mers are instead owing to point mutations in high-identity sequences, there should be morph-specific *k*-mers in both morphs.

Most (>98%) of the mapped *k*-mers in the A versus O and A versus I comparisons aligned perfectly to a single -1.5 mb region of the unlocalized scaffold 2 of chromosome 13, in the A-morph assembly (Fig. 3a and Extended Data Table 1). This is the same region of the A-morph assembly that was previously identified in the standard GWAS (Fig. 2). In contrast, only ~0.3% of the associated *k*-mers in the A versus O comparison were found anywhere in the O assembly and, similarly, only ~0.2% of the significant *k*-mers in the A versus I analysis mapped to the I assembly (Extended Data Table 1). These results thus suggested that a large region of genomic content is unique to the A haplotype.

Given that A and I females share their immature colour pattern<sup>29,36</sup>, we then tested for *k*-mer associations that would distinguish both A and I females from O females and found 85,134 such *k*-mers (Extended Data Table 1). When mapped to the A assembly, a majority of these *k*-mers were found near the start of the unlocalized scaffold 2 of chromosome 13 (Fig. 3a), where we previously reported pronounced divergence of O females (Fig. 2b,c). However, when mapped to the I assembly, most

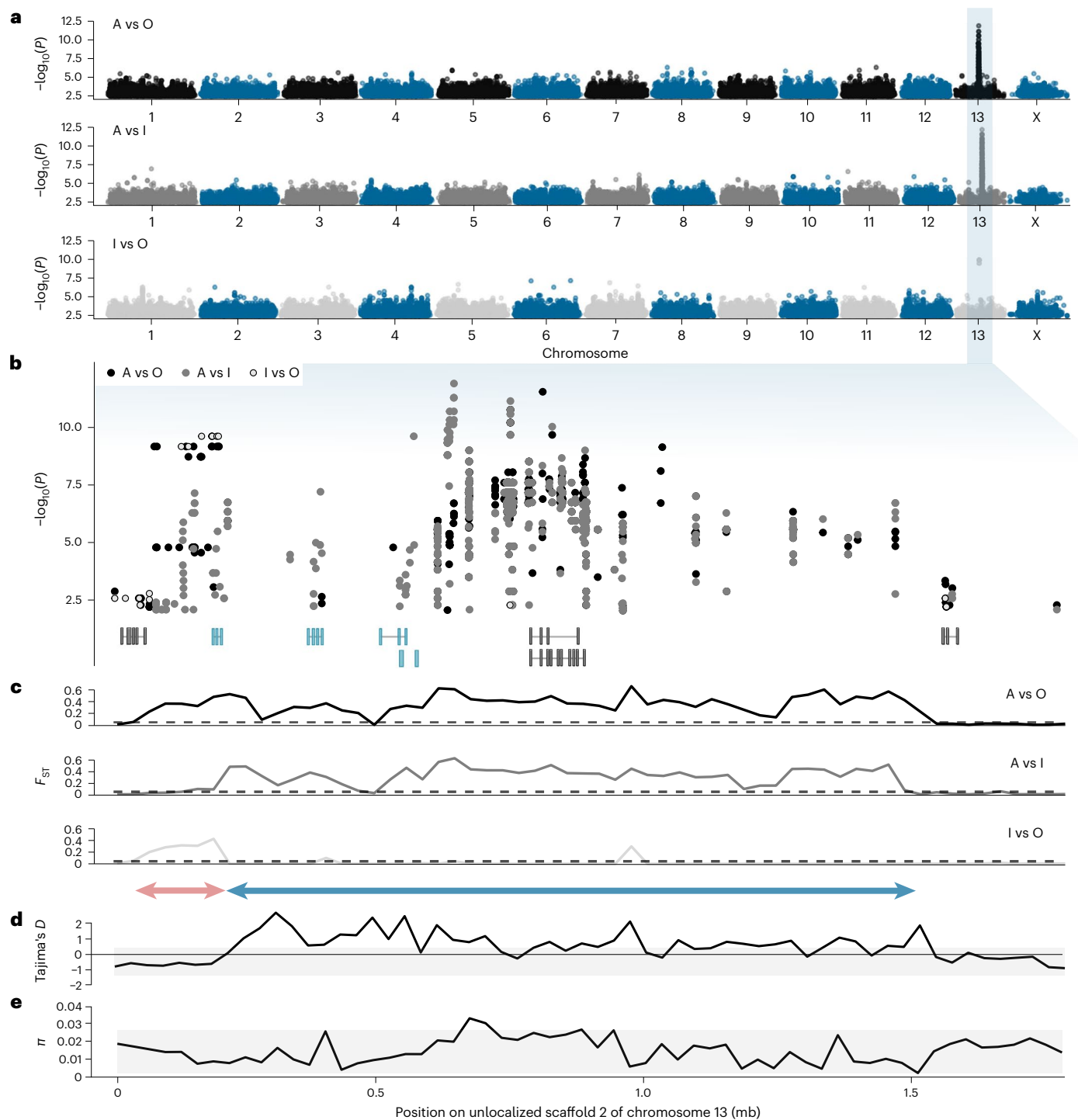
of the significant *k*-mers were found in a different region of the same scaffold, separated by approximately 3.5 mb (Fig. 3b). These results thus suggested that A and I females share genomic content that is absent in O. However, in the I haplotype this content occupies a different chromosomal location.

To further investigate the distribution of genomic content among morphs, we plotted the standardized number of mapped reads (read depths) along the -1.5 mb region of the A assembly that included most of the significant *k*-mers (Extended Data Fig. 1). Here, we expected read depth values around 0.5 (heterozygous) or 1.0 (homozygous) for all A samples, whereas I and O samples should have read depths of 0, if genomic content is uniquely present in the A allele (because I and O individuals lack the A allele; Fig. 1). Read depths confirmed that male-mimicking A females are differentiated by genomic content. Specifically, there are two windows of the A assembly (of ~400 kb and ~500 kb) where no I or O data maps to the assembly after filtering repetitive sequences (Fig. 3c), and that are therefore uniquely present in A females. These two windows of A-specific content are separated by a region between -0.6 and -1.0 mb that is shared among all morphs (Fig. 3c), and highly divergent in SNP-based comparisons involving the A morph (Fig. 2b). Finally, the region including most significant *k*-mers in the A and I versus O comparison is present in all A and I samples but absent in all O samples, except for one individual (Fig. 3c and Supplementary Text 3). As noted in the *k*-mer GWAS, this region of genomic content shared by A and I individuals is located in different regions, separated by ~3.5 mb, in the two assemblies (Fig. 3d).

By combining reference-based GWAS, reference-free GWAS and read-depth approaches, we have identified three haplotypes controlling morph development in the common bluetail. The A and I haplotypes share ~150 kb that are absent in O. The A haplotype has two additional windows of unique genomic content, adding up to ~900 kb. In the A haplotype, a single -1.5 mb window (hereafter the morph locus) thus contains the regions of unique genomic content, the region exclusively shared between A and I, and the SNP-rich region present in all morphs. In the I haplotype the region exclusively shared with A occupies a single and different locus separated by about 3.5 mb (Fig. 4a). These large and compounded differences in genomic content between haplotypes suggest that multiple structural changes on a multi-million base-pair region were responsible for the evolution of novel female morphs in *Ischnura* damselflies.

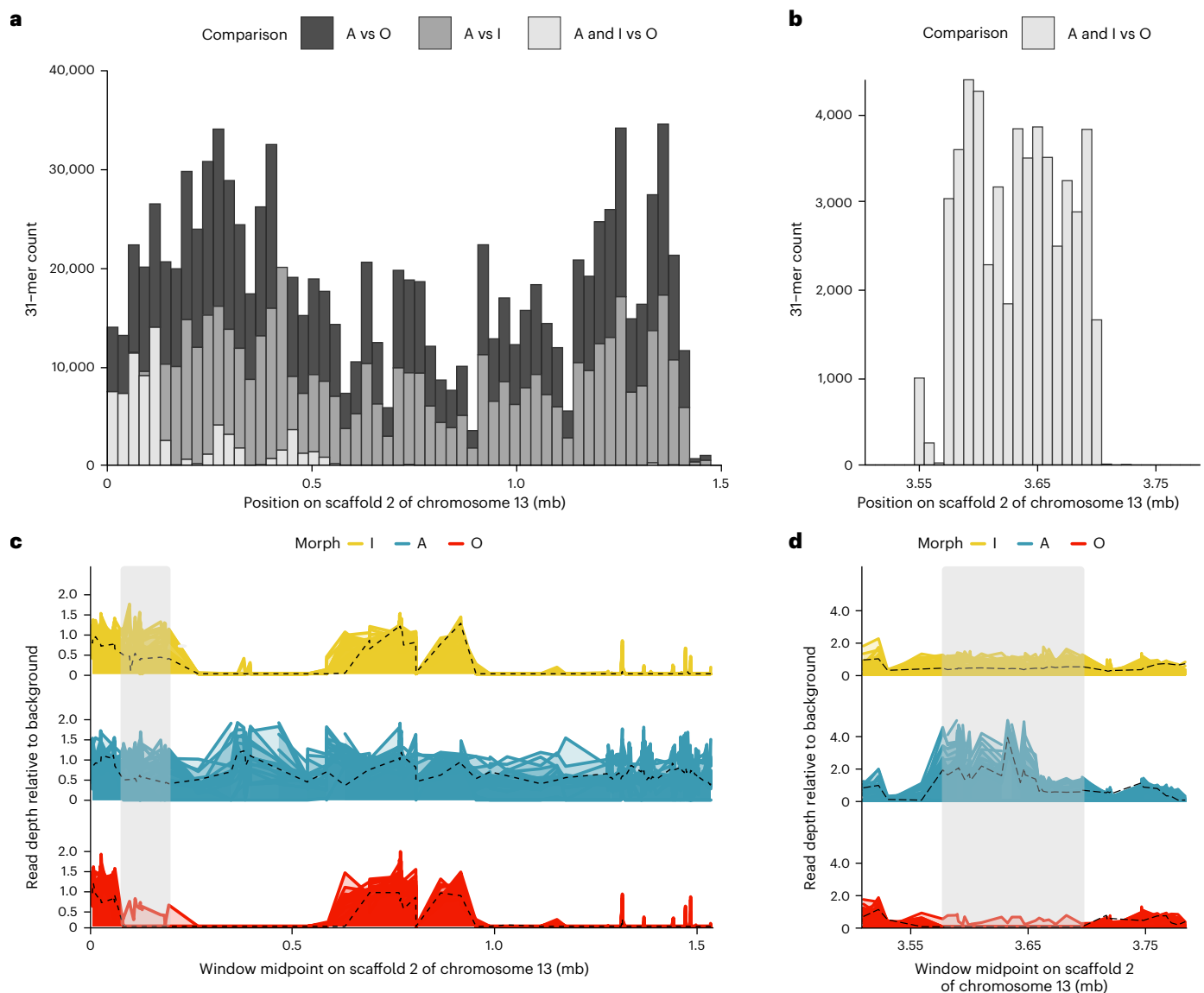
### TE propagation and recombination probably explain the origins of novel female morphs

Based on previous inferences of the historical order in which female morphs evolved (Fig. 1), we hypothesized that genomic divergence first occurred between O and A females, with some genomic content being then translocated from A into an O background, leading to the evolutionary origin of I females. We analysed SVs between morphs to test this hypothesis (Extended Data Fig. 1 and Supplementary Text 4) and uncovered evidence of a ~20 kb sequence in the O haplotype that is duplicated and inverted in tandem in derived morphs (A and I; Fig. 4b and Extended Data Fig. 2). An investigation of the reads mapping to the inversion breakpoints suggested that additional duplications in the A genome, presumably via TE proliferation, may be related to the evolution of inter-sexual mimicry (Fig. 4b and Extended Data Fig. 3). Interestingly, TE content is enriched and recombination is reduced not just in the vicinity of the morph locus, but across the entire chromosome 13 (Extended Data Figs. 4 and 5, and Supplementary Text 4). Finally, evidence of a translocation of an A-derived genomic region back into an O background (Extended Data Fig. 6 and Supplementary Text 4) implied that the I morph evolved from an ectopic recombination event between A and O morphs (Fig. 4b). This scenario is also consistent with our previous *k*-mer GWAS and read-depth results, where we found that the only region differentiating both A and I females from O females is located ~3.5 mb in the I haplotype.



**Fig. 2 | Morph determination in *I. elegans* is controlled in a -1.5 mb region of chromosome 13.** **a**, SNP-based GWAS in all pairwise analyses between morphs. Genomic DNA from 19 wild-caught females of each colour morph and of unknown genotype was extracted and sequenced for these analyses. Illumina short reads were aligned against an A morph genome assembly, generated from nanopore long-read data (Extended Data Fig. 1). **b**, A closer look at the SNP associations on the unlocalized scaffold 2 of chromosome 13, which contained all highly significant SNPs. Transcripts expressed in at least one adult of both *I. elegans* and *I. senegalensis* are shown at the bottom (see also Fig. 6). Grey transcripts are shared by all morphs, whereas blue transcripts are uniquely present in A or A and I samples (see ‘Shared and morph-specific

genes reside in the morph locus’). The y axis in **a** and **b** indicates unadjusted  $-\log_{10}P$  values calculated from chi-squared tests. **c**,  $F_{ST}$  values averaged across 30 kb windows for the same pairwise comparisons as in the SNP-based GWAS. The dashed line marks the 95th percentile of all non-zero  $F_{ST}$  values across the entire genome. The red double arrow shows the region of elevated divergence between O and both A and I samples (~50 kb–0.2 mb). The blue double arrow shows the region of elevated divergence between A and both O and I samples (~0.2 mb–1.5 mb). **d,e**, Population-level estimates of Tajima's  $D$  (**d**) and  $\pi$  (**e**) averaged across 30 kb windows. The shaded area contains the 5th–95th percentile of all genome-wide estimates.



**Fig. 3 | Female morphs of *I. elegans* differ in genomic content.** **a, b**, Number of significant *k*-mers (below the 5% false-positive threshold; Methods) associated with pairwise genome-wide analyses and mapped to the unlocalized scaffold 2 of chromosome 13, in the A-morph assembly (**a**) and the I-morph assembly (**b**). **c, d**, Standardized read depths along the unlocalized scaffold 2 of chromosome 13, relative to background coverage of the A-morph assembly (**c**) and the I-morph

assembly (**d**). Solid lines (yellow, blue and red) show short-read data (19 samples per morph) and black dashed lines represent long-read data (1 sample per morph). Grey areas show regions of genomic content present in A and I individuals, but absent in all but one O sample. Note that different regions of the scaffold are plotted for the two assemblies (see main text).

### Male mimicry is a trans-species polymorphism

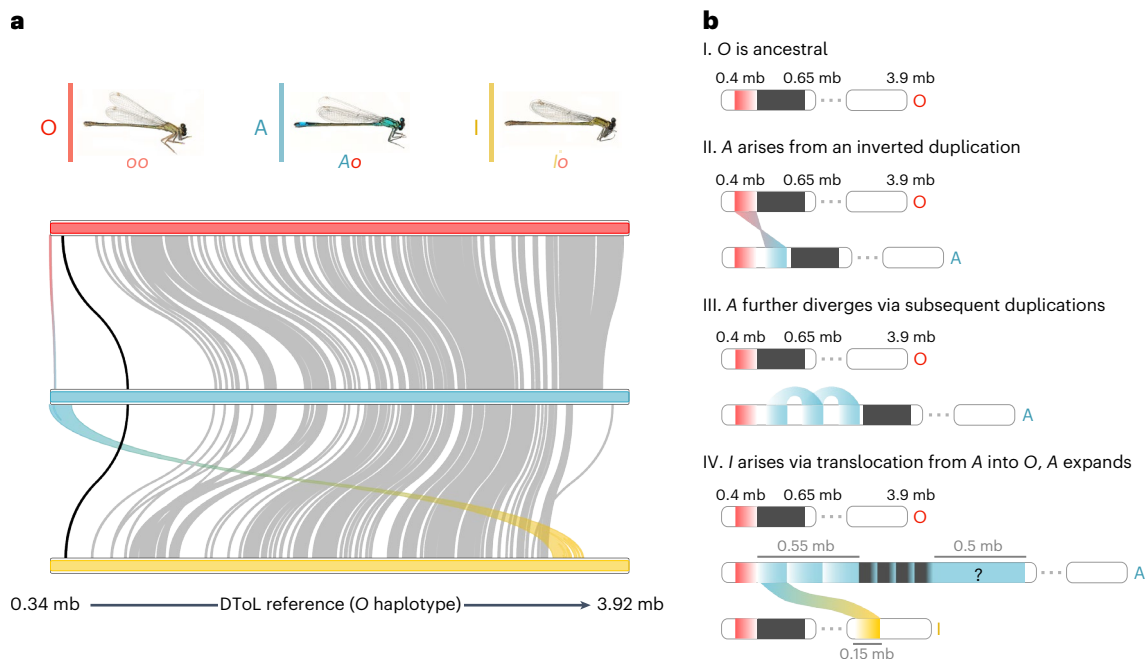
Ancestral state reconstruction of female colour states had previously pointed to an ancient origin of male mimicry in the clade that includes *I. elegans* and several other widely distributed *Ischnura* damselflies<sup>28</sup> (Fig. 1). We investigated whether male mimicry is in fact a trans-species polymorphism using de novo genome assemblies from the closely related tropical bluetail (*I. senegalensis*; Extended Data Fig. 1). *I. senegalensis* shares a common ancestor with *I. elegans* about 5 Myr ago<sup>28</sup>, and has both a male-mimicking A morph and a non-mimicking morph, which resembles the O females of *I. elegans*<sup>28,37</sup> (Fig. 5a).

We reasoned that if morph divergence is ancestral, the genomic content that is uniquely present in A females or shared by A and I females in *I. elegans* should be at least partly present in A females of *I. senegalensis*, but absent in the alternative O-like female morph (Supplementary Text 5). This prediction was supported by differences in standardized read depths between the A and O-like pool of

*I. senegalensis*, specifically at the morph locus of *I. elegans* (Fig. 5b and Supplementary Text 5). A shared genomic basis of inter-sexual mimicry for the two species was also supported by the same ~20 kb inversion signature in the A pool against an O assembly, as detected in A and I females of *I. elegans* (Extended Data Fig. 7). Finally, assembly alignments between O-like and A haplotypes of *I. senegalensis* showed that the A-specific genomic region of *I. elegans* is partly present in the A but not the O-like assembly of *I. senegalensis* (Fig. 5c).

### Shared and morph-specific genes reside in the morph locus

Finally, we examined gene content and expression patterns in the morph locus. As female morphs differ in genomic content as well as sequence, the phenotypic effects of the morph locus could come about in at least three non-exclusive ways. First, entire gene models may be present in some morphs and absent in others. Second, genes present in all morphs may differ in expression patterns. Third, genes may encode



**Fig. 4 | SVs differentiate morph haplotypes in the common bluetail damselfly (*I. elegans*).** **a**, Alignment between morph-specific genomes assembled from long-read nanopore samples with genotypes *Ao*, *Io* and *oo*. Grey lines represent alignments of at least 5 kb and >70% identity. The black line connects regions of genomic content shared by the three morphs within the morph locus. The red to blue gradient represents a ~20 kb region that carries an inversion signature in *A* and *I* females relative to the *O* haplotype (Extended Data Fig. 2). The blue to yellow gradient represents a ~150 kb alignment between the start of the unlocalized scaffold 2 of chromosome 13 in *A* and a region ~3.5 mb apart in the *I* haplotype. Coordinates at the bottom are based on the DTol reference assembly. **b**, Schematic illustration of the hypothetical sequence of events responsible for the evolution of novel female morphs. First, a sequence originally present in *O* was duplicated and inverted in tandem, potentially causing the initial divergence

of the *A* allele. Second, part of this inversion was subsequently duplicated in *A*, in association with a putative TE, leading to multiple inversion signatures in the *A* haplotype relative to an *O* reference (Extended Data Fig. 3). Finally, part of the *A* duplications were translocated into a position ~3.5 mb downstream into an *O* background, giving rise to the *I* allele. Currently, *A* females are also characterized by another region of unique content and unknown origin (question mark). *A* females show elevated sequence divergence in the internal region of the morph locus that is shared by all haplotypes (dark grey bars; see also black line in **a**). Coordinates on the *O* haplotype are based on the (DTol) reference assembly. Grey numbers in IV give the approximate size of genomic sequences in *A* and *I* that are absent in *O*. Damselfly images adapted from ref. 25 under a Creative Commons licence CC BY 4.0.

different amino acid sequences in different female morphs. We used newly generated and previously published<sup>38</sup> RNA sequencing (RNA-seq) data to investigate these questions (Extended Data Fig. 1), and capitalized on the annotations of the reference genome of *I. elegans*<sup>34</sup>, as well as transcripts assembled de novo in our *A*-morph genome assembly. Because the genetic basis of inter-sexual mimicry is shared between *I. elegans* and *I. senegalensis* (Fig. 5), we focus on genes that are expressed in both species in at least one individual (Fig. 6a).

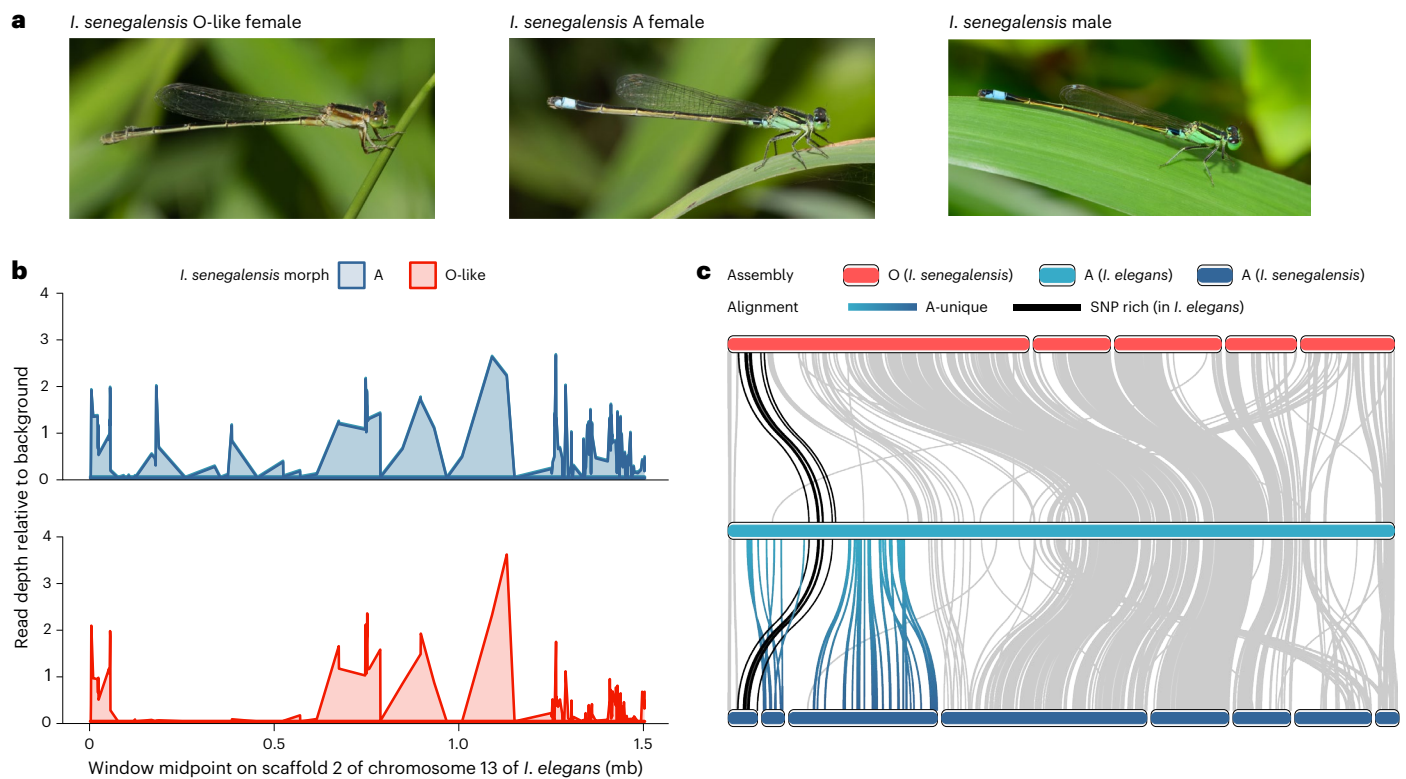
Three transcripts (from two predicted genes) in the morph locus are expressed in *A* females of *I. senegalensis*, and in *A* and *I* females of *I. elegans*, but never in *O* or *O*-like females (Fig. 6b). Only one of these gene models (Afem.4094) could be functionally annotated, and appears to encode a long interspaced nuclear element (LINE) retrotransposon in the clade Jockey (Supplementary Text 6). This gene also exhibited expression changes in *I* females that reflect their colour development trajectory of initial resemblance to *A* females, followed by an overall appearance similar to *O* females upon sexual maturation (Supplementary Text 6). Notably, RepeatModeler and RepeatMasker detected signatures of the Jockey family at the same locus as the mapping locations of the *A* reads that had suggested a propagation of TEs in our SV analyses (Fig. 6a and Extended Data Fig. 3). Thus, these results further support that TEs are responsible for the evolution and expansion of the male-mimicry allele.

We also identified three gene models that are shared by all haplotypes and expressed in both species. The three predicted genes encode zinc-finger domain proteins (Fig. 6b and Supplementary Text 6), which

are known to participate in transcriptional regulation<sup>39</sup>. However, we found no conclusive evidence of differential expression, nor evidence of non-synonymous substitutions between morphs shared by both *I. elegans* and *I. senegalensis* (Supplementary Text 6). While we see genes of a potentially regulatory function reside in the morph locus, understanding their role in morph development will probably require higher temporal and spatial resolution of gene expression data.

## Discussion

Sexual dimorphism, where males and females have markedly distinct colour patterns, has led to multiple evolutionary origins of female-limited polymorphisms and potential male mimicry in *Ischnura* damselflies<sup>28</sup>. Here, we present a genomic glance into how these morphs evolve, setting the stage for future functional work to unravel the reversal of sexual phenotypes in damselfly sexual mimicry. Male mimicry in the common bluetail is controlled by a single genomic region in chromosome 13 (Figs. 2 and 3). Our data suggest that this morph locus probably evolved with the accumulation of novel and potentially TE-derived sequences in the male mimicry haplotype (Fig. 4), which is shared by male-mimicking females of species diverging more than 5 Myr ago (Fig. 5). More recently, a rare recombination event involving part of the novel *A* genomic content has triggered the origin of a third female morph (Fig. 4), which shares its sexually immature colouration and patterning with *A* females, and shares its sexually mature overall appearance with *O* females<sup>27</sup>. The morph locus contains a handful of genes, some of which may have evolved with TE propagation in the *A*



**Fig. 5 | A shared genomic basis of A females in *I. elegans* and *I. senegalensis*.** **a**, *I. senegalensis* is a female-dimorphic species, where one female morph (O-like) is distinctly different from males and resembles O females in *I. elegans*, and the other female morph (A) is a male mimic. Photo credit: Mike Hooper. **b**, Standardized read depth of pool-seq samples ( $n = 30$  females of each morph per pool) of *I. senegalensis*, against the A-morph assembly of *I. elegans*, calculated in 500 bp windows. The x-axis shows the first 1.5 mb of the unlocalized scaffold 2 of chromosome 13. **c**, Alignments between morph-specific genomes

from a homozygous O-like female of *I. senegalensis* (top), an Ao female of *I. elegans* (middle) and a homozygous A female of *I. senegalensis* (bottom). Lines connecting the assemblies represent alignments of at least 500 bp and >70% identity. The black line connects genomic content in the morph locus, which is shared by the three morphs of *I. elegans*. In *I. elegans*, this region is rich in SNPs differentiating A females from the other two morphs (see Fig. 2b). The blue–turquoise gradient connects sequences uniquely present in the A morphs of *I. elegans* and *I. senegalensis*.

haplotype, and are therefore absent from O individuals (Fig. 6). However, existing annotations provide only a hint on how these genes may influence morph development. Our results thus echo recent calls for a broader application of functional validation tools, in order to understand how lineage-specific genes contribute to phenotypic variation in natural populations<sup>40</sup>.

This study underscores two increasingly recognized insights in evolutionary genomics. First, there is mounting evidence that SVs abound in natural populations and often underpin complex and ecologically relevant phenotypic variation<sup>41</sup>, such as discrete phenotypic polymorphisms<sup>10,13,15,20</sup>. Nonetheless, traditional GWAS approaches based on SNPs can easily miss SVs, as these approaches are contingent on the genomic content of the reference assembly<sup>42</sup>. Among other novel approaches to tackle this problem<sup>42</sup>, a reference-free  $k$ -mer-based GWAS, as implemented here, is a powerful method to identify variation in genomic content and sequence, especially when the genomic architecture of the trait of interest is initially unknown<sup>35</sup>. In this study, we did not know a priori which of the three morphs, if any, would harbour unique genomic content. Had we ignored differences in genomic content between morphs and based our GWAS analysis solely on the DTOL (O) reference assembly, we would have failed to identify SNPs between I and O morphs (Extended Data Fig. 8), and the origin of I females via a translocation of A content would have been obscured.

Second, a role for TEs in creating novel and even adaptive phenotypic variation is increasingly being recognized<sup>43,44</sup>. Here, we found that a ~400 kb region of unique genomic content, possibly driven by LINE transposition, is associated with the male-mimicry phenotype in

at least two species of *Ischnura* damselflies. TE activity can contribute to phenotypic evolution by multiple mechanisms. For instance, TEs may modify the regulatory environment of genes in their vicinity, by altering methylation<sup>45</sup> and chromatin conformation patterns<sup>46</sup>, or by providing novel *cis*-regulatory elements<sup>47</sup>. The male-mimicry region in *I. elegans* is located between two coding genes with putative DNA-binding domains, and that may thus act as transcription factors. However, our expression data do not provide unequivocal support for differential regulation of either of these genes between female morphs. Importantly, currently available expression data come from adult specimens, as female morphs are not visually discernible in aquatic nymphs. Yet, the key developmental differences that produce the adult morphs are probably directed by regulatory variation during earlier developmental stages. Now that the morph locus has been identified, future work can address differential gene expression at more relevant developmental stages, before colour differences between morphs become apparent.

TEs can also contribute to phenotypic evolution if they become domesticated, for example, when TE-encoded proteins are remodelled through evolutionary change to perform adaptive host functions<sup>48</sup>. We found two transcripts located in A-specific or A/I-specific regions that are probably derived from LINE retrotransposons and are actively expressed in the genomes that harbour them (Fig. 6b). It is therefore possible that these transcripts participate in the development of adult colour patterns, which are initially more similar between A and I females than between either of these morphs and O females<sup>27,29</sup>. Yet, functional work on these transcripts is required to ascertain their role in morph



**Fig. 6 | The morph locus of *I. elegans* is situated in the unlocalized scaffold 2 of chromosome 13. a**, Diagram of the -1.5 mb morph locus on the A-morph assembly, showing from top to bottom: morph-specific read depth coverage; the location of LINE retrotransposons in the the Jockey family; the mapping locations of A-derived reads with a previously detected inversion signature against O females; and transcripts expressed in at least one adult individual of both *I. elegans* and *I. senegalensis*. Transcripts plotted in black are present in both the A and O assemblies, while transcripts in blue are located in genomic regions that are unique to the A haplotype or are shared between A and I but not the O allele.

**b**, Functional annotations and sex- and morph-specific expression of transcripts. Square fill indicates whether transcript expression was detected in each group. RNA-seq data for *I. elegans* comes from whole-thorax samples from sexually immature and sexually mature wild-caught adults ( $n = 3$  females of each morph and 3 males). RNA-seq data for *I. senegalensis* comes from a recent study in which the abdomen, head, thorax and wings were sampled in two females of each morph and two males (one individual of each group sampled upon emergence and one sampled after two days).

determination. Finally, TEs can become sources of novel small RNAs that play important regulatory roles<sup>49</sup>, including in insect sex determination<sup>50</sup>. Thus, future work should also address non-coding RNA expression and function in the morph locus.

Our results also provide molecular evidence for previous insights, gained by alternative research approaches, on the micro- and macroevolution of female-limited colour polymorphisms. A wealth of population data in southern Sweden has shown that female-morph frequencies are maintained by balancing selection, as they fluctuate less than expected due to genetic drift<sup>31</sup>. Behavioural and field experimental studies indicate that such balancing selection on female morphs is mediated by negative frequency-dependent male harassment<sup>51,52</sup>. We add to these earlier results by showing a molecular signature consistent with balancing selection in the genomic region where A females differ from both of the non-mimicking morphs. Sexual conflict is expected

to have profound effects on genome evolution, but there are few examples of sexually antagonistic traits with a known genetic basis, in which predictions about these genomic effects can be tested<sup>53,54</sup>. Here, the signature of balancing selection on the morph locus matches the expectation of inter-sexual conflict resulting in negative frequency-dependent selection and maintaining alternative morph alleles over long periods.

Similarly, a recent comparative study based on phenotypic and phylogenetic data inferred a single evolutionary origin of the male-mimicking morph shared by *I. elegans* and *I. senegalensis*<sup>28</sup>. Our present results strongly support this common origin. This is because A females in both species share unique genomic content, including associated transcripts, and an inversion signature against the ancestral O morph (Fig. 5 and Extended Data Fig. 7). Nonetheless, these data are consistent with both an ancestral polymorphism and ancestral introgression being responsible for the spread of male mimicry



across the clade. A potential role for introgression in the evolution of male mimicry is also suggested by rampant hybridization between *I. elegans* and its closest relatives<sup>55</sup>, and by the fact that *I. elegans* and *I. senegalensis* can hybridize millions of years after their divergence, at least in laboratory settings<sup>56</sup>. The identification of the morph locus in *I. elegans* enables future comparative genomics studies to disentangle the relative roles of long-term balancing selection and introgression in shaping the widespread phylogenetic distribution of female-limited polymorphisms in *Ischnura* damselflies.

Finally, our results open up new lines of enquiry on how the genomic architecture and chromosomal context of the female polymorphism may influence its evolutionary dynamics. Our data are consistent with the evolution of a third morph due to an ectopic recombination event that translocated genomic content from the A haplotype back into an O background. Ectopic recombination can occur when TE propagation generates homologous regions in different genomic locations<sup>57,58</sup>, and may be facilitated by the excess of TE content in chromosome 13 (Extended Data Fig. 4). The male reproductive morphs in the ruff (*Calidris pugnax*) are one of few previous examples of a novel phenotypic morph arising via recombination between two pre-existing morph haplotypes<sup>10</sup>. In pond damselflies, female polymorphisms with three or more female morphs are not uncommon, and in some cases female morphs exhibit graded resemblance to males<sup>59</sup>. It is therefore possible that recombination, even if rare, has repeatedly generated diversity in damselfly female morphs.

While recombination might have had a role in generating the novel I morph, we observe reduced recombination over the morph locus in comparison to the rest of the genome of *I. elegans* (Extended Data Fig. 5). However, this reduction in recombination is not limited to the morph locus and its vicinity, but rather pervasive across chromosome 13 (Extended Data Fig. 5). This unexpected finding suggests two alternative causal scenarios. First, selection for reduced recombination at the morph locus, following the origin of sexual mimicry, could have spilled over chromosome 13, facilitating a subsequent accumulation of TEs and further reducing recombination<sup>60</sup>. Second, TE enrichment and reduced recombination may have preceded the evolution of female morphs, and facilitated the establishment and maintenance of the female polymorphism by balancing selection.

Both historical scenarios are compatible with a morph locus reminiscent of a supergene, which is defined by tight genetic linkage of multiple functional loci<sup>61</sup>. However, an alternative and parsimonious explanation is that the novel sequences in A and I females and their flanking genes may not code for anything important for the male-mimicking phenotype as such, but simply disrupt a region of chromosome 13 that is required for the development of ancestral sexual differentiation. The observation that I females carry part of the sequence that originated in A in a different location of the scaffold (Fig. 4b), and still develop some male-like characters (for example, black thoracic stripes), could come about if insertions anywhere on a larger chromosomal region disrupt female suppression of the male phenotype, although with variable efficacy depending on the exact location or insertion size.

### Concluding remarks

Recent years have witnessed an explosion of studies uncovering the loci behind complex phenotypic polymorphisms in various species. An emerging outlook is that not all polymorphisms are created equal, with some governed by massive chromosomal rearrangements<sup>15–17</sup>, and others by a handful of regulatory sites<sup>11,12,18</sup>. Our results contribute to this growing field by uncovering a single causal locus that features structural variation and morph-specific transcripts in the female-limited morphs of *Ischnura* damselflies. These morphs not only differ in numerous morphological and life-history traits<sup>32,62,63</sup> and gene expression profiles<sup>24,25</sup>, but they include a male mimic that is maintained by balancing frequency-dependent selection.

Our findings enable future studies on the developmental basis of such male mimicry, with consequences for a broader understanding of the evolutionary dynamics of sexual dimorphism and the cross-sexual transfer of trait expression<sup>64</sup>.

## Methods

### *I. elegans* samples

Samples for morph-specific genome assemblies of *I. elegans* were obtained from F<sub>1</sub> individuals with genotypes *Ao*, *Io* and *oo* (one adult female of each genotype). In June 2019, recently mated O females were captured in field populations in southern Sweden. These females oviposited in the lab within 48 h, and their eggs were then released into outdoor cattle tanks seeded with *Daphnia* and covered with synthetic mesh. Larvae thus developed under normal field conditions and emerged as adults during the summer of 2020. Emerging females were kept in outdoor enclosures until completion of adult colour development<sup>25,65</sup>. Fully mature females were phenotyped, collected in liquid nitrogen and kept at  $-80^{\circ}\text{C}$ . Because all of these females carry a copy of the most recessive allele *o*, individuals of the A and I morph are heterozygous, with genotypes *Ao* and *Io*, respectively.

A total of 19 resequencing samples of each female morph of *I. elegans* were also collected from local populations in southern Sweden, within a  $40 \times 40$  km area (Supplementary Table 1). Samples were submerged in 95% ethanol and stored in a  $-20^{\circ}\text{C}$  freezer until extraction. Additionally, 24 individuals (six adult females of each morph and six males) were collected for RNA-seq analysis in a natural field population (Bunkeflostrand) in southern Sweden, in early July 2019. These samples were transported on carbonated ice and stored in  $-80^{\circ}\text{C}$  until extraction.

### *I. senegalensis* samples

Adults of *I. senegalensis* (30 adult females of each morph) were collected for pool sequencing from a population on Okinawa Island in Japan ( $26.148^{\circ}\text{N}$ ,  $127.795^{\circ}\text{E}$ ) in May 2016. Samples were visually determined to sex and morph and stored in 99% ethanol until extraction. Samples for morph-specific genome assemblies of *I. senegalensis* were obtained from a population in Clementi Forest, Singapore ( $1.33^{\circ}\text{N}$ ,  $103.78^{\circ}\text{E}$ ). Because the A allele is recessive in *I. senegalensis*, all females with the A phenotype are homozygous. To obtain a homozygous O-like sample, we developed primers (forward: CGCGTATGATGGTCCGA, reverse: GGCTGCTTACACCAATGCAA) for an A-specific sequence that is shared by A females of the two species (318,131–318,213 bp on the A haplotype of *I. elegans*). We used the mapped pool-seq data to identify fixed SNPs between species and tailor the primer sequences accordingly. We then tested the primers in 20 A females of *I. senegalensis* using a 328 bp fragment of the histone H3 gene (forward: ATGGCTCGTACCAAGCAGACGGC, reverse: ATATCCTTGGGCATGATGGTGAC)<sup>66</sup> as a positive control for the polymerase chain reaction. Once validated, we utilized these primers to identify O-like females lacking the A allele and selected one of these samples for whole genome sequencing.

### DNA extraction, library preparation and sequencing

High molecular weight (HMW) DNA was extracted from one *I. elegans* female of each genotype (*Ao*, *Io*, *oo*), using the Nanobind Tissue Big Extraction Kit (NB-900-701-01, Circulomics Inc. (PacBio)). HMW DNA was isolated from homozygous females of each morph of *I. senegalensis*, using the Monarch HMW DNA Extraction Kit for Tissue (T3060S, New England BioLabs Inc.). DNA from resequencing samples was isolated using either a modified protocol for the DNeasy Blood and Tissue Kit (19053, Qiagen) or the KingFisher Cell and Tissue DNA Kit (Cat no. N11997, ThermoFisher Scientific). *I. senegalensis* DNA was extracted from muscle tissues in thoraxes using Maxwell 16 LEV Plant DNA Kit (AS1420, Promega). Details on extraction and library preparation protocols are provided in Supplementary Text 1.

Sequencing libraries were constructed from each HMW DNA sample for the Nanopore LSK-110 ligation kit (Oxford Nanopore Technologies). Adapter ligation and sequencing of *I. elegans* samples were carried out at the Uppsala Genome Centre, hosted by SciLife Lab. Each sample was sequenced on a PromethION R10.4 with one nuclease wash and two library loadings. Library preparation and sequencing of *I. senegalensis* samples were carried out by the Integrated Genomics Platform, Genome Institute of Singapore, A-STAR, Singapore. Each sample was sequenced on a PromethION R9.4 flow cell, with two nuclease washes and three library loadings.

### RNA extraction and sequencing

Whole-thorax samples were ground into a fine powder using a Tissue-Lyser and used as input for the Spectrum Plant Total RNA Kit (STRN50, Sigma Aldrich), including DNase I treatment (DNASE10, Sigma Aldrich). Library preparation and sequencing were performed by SciLife Lab at the Uppsala Genome Centre. Sequencing libraries were prepared from 300 ng of RNA, using the TrueSeq stranded mRNA library preparation kit (20020595, Illumina Inc.) including polyA selection and unique dual indexing (20022371, Illumina Inc.), according to the manufacturer's protocol. Sequencing was performed on the Illumina NovaSeq 6000 SP flowcell with paired-end reads of 150 bp.

### De novo genome assembly

Bases in raw Oxford Nanopore Technologies reads from *I. elegans* were called using Guppy v.4.0.11 (*Ao* and *Io* data) and Guppy v.5.0.11 (*oo* data) (<https://nanoporetech.com/>). Low-quality reads ( $q$ -score <7 for v.4.0.11 and <10 for v.5.0.11) were subsequently discarded. High quality reads were assembled using the Shasta long-read assembler v.0.7.0<sup>67</sup>. Each assembly was conducted under four different configuration schemes, which modified the June 2020 nanopore configuration file (<https://github.com/chanzuckerberg/shasta/blob/master/conf/Nanopore-Jun2020.conf>) in alternative ways (Supplementary Table 3). Assembly metrics were compared among Shasta configurations for each morph using AsmQC<sup>68</sup> (<https://sourceforge.net/projects/amos/>) and the stats.sh script in the BBTools suite (<https://sourceforge.net/projects/bbmap>). The assembly with greater contiguity (that is, highest contig N50, highest average contig length and highest percentage of the main genome in scaffolds >50 kb) was selected for polishing and downstream analyses.

Bases in raw Oxford Nanopore Technologies reads from *I. senegalensis* samples were called using Guppy v.6.1.5. Reads with quality score <7 were subsequently discarded. High quality reads were assembled using the Shasta long-read assembler v.0.7.0<sup>67</sup> and the configuration file T2 (Supplementary Table 3), which was also selected for the *Io* and *oo* assemblies of *I. elegans*.

Morph-specific assemblies of *I. elegans* were first polished using the Oxford Nanopore Technologies reads mapped back to their respective assembly with minimap2 v.2.22-r1110<sup>69</sup>, and the PEPPER-Margin-DeepVariant pipeline r0.4<sup>70</sup>. Alternative haplotypes were subsequently filtered using purge\_dups v.0.0.3<sup>71</sup>, to produce a single haploid genome assembly for each sample. The *I. elegans* draft assemblies were then polished with short read data from one resequencing sample (TE-2564-SwD172\_S37; Supplementary Table 1), using the POLCA tool in MaSuRCA v.4.0.4<sup>72</sup>. For every draft and final assembly of *I. elegans*, we computed quality metrics as mentioned above and assessed the completeness of conserved insect genes using BUSCO v.5.0.0<sup>73</sup> and the 'insecta\_odb10' database (Supplementary Fig. 1). For *I. senegalensis*, we report quality metrics of the final assemblies (Supplementary Fig. 2).

### Scaffolding with the DTOL super assembly

During the course of this study, a chromosome-level genome of *I. elegans* was assembled by the DTOL Project<sup>34</sup>, based on long-read (PacBio) and short-read (Illumina) data, as well as Hi-C (Illumina)

chromatin interaction data. Of the total length of this assembly, 99.5% is distributed across 14 chromosomes, one of which (no. 13) is fragmented and divided into a main assembly and five unlocalized scaffolds.

We used RagTag v.2.10<sup>74</sup> to scaffold each of our morph-specific assemblies based on the DTOL reference (Supplementary Text 2). Scaffolding was conducted using the nucmer v.4.0.0<sup>75</sup> aligner and default RagTag options. Morph-specific scaffolded genomes were also aligned to each other using nucmer and a minimum cluster length of 100 bp. Alignments were then filtered to preserve only the longest alignments in both reference and query sequences, and alignments of at least 5 kb. These assembly alignments were then used to visualize synteny patterns across morphs, in the region uncovered in our association analyses (Extended Data Fig. 1), using the package RIdeogram v.0.2.2<sup>76</sup> in R v.4.2.2<sup>77</sup>.

### Reference-based (SNP) GWAS

We first investigated genomic divergence between morphs using a standard GWAS approach based on SNPs (Extended Data Fig. 1). Initially, we conducted preliminary analyses using different morph assemblies as mapping reference. Once the A-specific genomic region was confirmed, we designated the A assembly as the mapping reference for the main analyses. Short-read data were mapped using bwa-mem v.0.7.17<sup>78</sup>. Optical and polymerase chain reaction duplicates were then flagged in the unfiltered bam files using GATK v.4.2.0.0<sup>79</sup>. Variant calling, filtering and sorting were conducted using bcftools v.1.12<sup>80</sup>, excluding the flagged reads. We retained only variant sites with mapping quality >20, genotype quality >30 and minor allele frequency >0.02 (that is, the variant is present in more than one sample). To avoid highly repetitive content, we filtered variants that had a combined depth across samples >1,360 (equivalent to all samples having ~50% higher than average coverage), and variants located in sites annotated as repetitive in either RepeatMasker v.1.0.93<sup>81</sup> or Red v.0.0.1<sup>82</sup>. The final variant calling file was analysed in pairwise comparisons (A versus O, A versus I, I versus O) using PLINK v.1.9<sup>83</sup> (<https://zzz.bwh.harvard.edu/plink/>). We report the  $-\log_{10}$  of  $P$  values for SNP associations in these pairwise comparisons.

### Reference-free ( $k$ -mer) GWAS

We created a list of all  $k$ -mers of length 31 in the short-read data (19 females per morph; Extended Data Fig. 1) following ref. 35, and counted  $k$ -mers in each sample using KMC v.3.1.0<sup>84</sup>. The  $k$ -mer list was filtered by the minor allele count;  $k$ -mers that appeared in less than five individuals were excluded.  $k$ -mers were also filtered by per cent canonized (that is, the per cent of samples for which the reverse complement of the  $k$ -mer was also present). If at least 20% of the samples including a given  $k$ -mer contained its canonized form, the  $k$ -mer was kept in the list. The  $k$ -mer list was then used to create a table recording the presence or absence of each  $k$ -mer in each sample. A kinship matrix for all samples was calculated from this  $k$ -mer table, and was converted to a PLINK<sup>83</sup> binary file, where the presence or absence of each  $k$ -mer is coded as two homozygous variants. In this step, we further filtered the  $k$ -mers with a minor allele frequency below 5%.

Because a single variant, be it an SNP or SV, will probably be captured by multiple  $k$ -mers, significance testing of  $k$ -mer associations requires a method to control for the non-independence of overlapping  $k$ -mers. We followed the approach developed by ref. 35, which uses a linear mixed model genome-wide association analysis implemented in GEMMA v.0.98.5<sup>85</sup>, and computes  $P$  value thresholds for associated  $k$ -mers based on phenotype permutations. We thus report  $k$ -mers below the 5% false-positive threshold as  $k$ -mers significantly associated with the female polymorphism in *I. elegans*. We conducted three  $k$ -mers based GWAS: (1) comparing male mimics to the putatively ancestral female morph (A versus O); (2) comparing male mimics to the most derived female morph (A versus I); and (3) comparing both derived female morphs (A and I) to the ancestral O females. For every analysis,

we then mapped the significant  $k$ -mers to all reference genomes using Blast v.2.22.28<sup>86</sup> for short sequences, and removed alignments that were below 100% identity and below full length. The mapped  $k$ -mers thus indicate the proportion of relevant genomic content present in each morph and how this content is distributed across each genome (Extended Data Table 1).

### Read-depth analysis

To validate the  $k$ -mer GWAS results of unique genomic content in A females relative to both I and O females, we plotted read depth across our region of interest (the unlocalized scaffold 2 of chromosome 13; see 'Results') in the A assembly (Extended Data Fig. 1). Short-read data (19 samples per morph) were mapped to the assembly with bwa-mem v.0.7.17<sup>78</sup> and reads with mapping score <20 were filtered, using Samtools v.1.14<sup>87</sup>. Long-read data (one sample per morph) were also mapped to the assembly using minimap2 v.2.22-r1110<sup>69</sup>, and quality filtering was conducted as above. Read depth was then averaged for each sample across 500 bp, non-overlapping windows using mosdepth v.0.2.8<sup>88</sup>. We also annotated repetitive content in the reference genome using RepeatMasker v.1.0.93<sup>81</sup> and Red v.0.0.1<sup>82</sup>, and filtered windows with more than 10% repetitive content under either method.

To account for differences in overall coverage between samples, we conducted the same procedure on a large (~15 mb) non-candidate region in chromosome 11 and calculated a 'background read depth' as the mean read depth across the non-repetitive windows of this region. We then expressed read depth in the candidate region as a proportion of the background read depth. Values around 1 thus indicate that a sample is homozygous for the presence of the sequence in a window. Values around 0.5 suggest that the sample only has one copy of this sequence in its diploid genome (that is, it is heterozygous). Finally, values of 0 imply that the 500 bp reference sequence is not present in the sample (that is, the window is part of an insertion or deletion).

We also investigated read-depth coverage on the I assembly, specifically across the region that was identified in the  $k$ -mer based GWAS as capturing content that differentiated both A and I females from O females (Fig. 3b and Extended Data Fig. 1). To do so, we followed the same strategy as above, except here we used a 15 mb region from chromosome 1 to estimate background read depth.

### Population genetics

We investigated the evolutionary consequences of morph divergence by estimating between-morph  $F_{ST}$  and population-wide Tajima's  $D$  and  $\pi$ . For these analyses, we used the A assembly as mapping reference and the same variant-calling approach as described for the SNP-based GWAS, but applied different filtering criteria (Extended Data Fig. 1). Specifically, invariant sites were retained and we only filtered sites with mapping quality score <20 and combined depth across samples >1,360 (equivalent to ~50% excess coverage in all samples).  $F_{ST}$  and  $\pi$  were estimated in pixy v.1.2.5<sup>89</sup> across 30 kb windows.  $F_{ST}$  was computed using the Hudson estimator<sup>90</sup>. Negative  $F_{ST}$  values were converted to zero for plotting. Tajima's  $D$  was estimated across 30 kb using vcftools v.0.1.17<sup>91</sup>. In all analyses, windows with >10% repetitive content according to either RepeatMasker v.1.0.93<sup>81</sup> or Red v.0.0.1<sup>82</sup> annotation were excluded.

### SVs

We used two complimentary approaches to identify SVs overlapping the genomic region uncovered by both  $k$ -mer-based and SNP-based GWAS. First, we mapped the raw data from each long-read sample to the assemblies of alternative morphs (for example *Ao* data mapped to *Io* and *oo* assemblies), and called SVs using Sniffles v.1.0.10<sup>92</sup> (Extended Data Fig. 1). These SV calls may represent fixed differences between morphs, within-morph polymorphisms or products of assembly error. We therefore used SamPlot v.1.3.0<sup>93</sup> and our short-read samples ( $n = 19$  per morph) to validate morph-specific SV calls (Extended Data Fig. 1).

Samplot identifies and plots reads with discordant alignments, which can result from specific types of SVs. For example, if Sniffles called a 10 kb deletion in the *Ao* and *Io* long-read samples relative to the *oo* assembly, we then constructed a Samplot for this region using short-read data, and expected to find support for such deletion in I and A samples, but not in O samples. We complemented this validation approach with a scan of the region of interest in each assembly, in windows of 250 and 500 kb, again using Samplot and the short-read data. If a SV appeared to be supported by the majority of short-read samples from an alternative morph, we zoomed in this SV and recorded the number of samples supporting the call in each morph.

### Linkage disequilibrium and TEs

To estimate linkage disequilibrium (LD), we used the same variant-calling file as for the SNP-based GWAS, which included only variant sites and was filtered by mapping quality, genotyping quality, minimum allele frequency and read depth, as described above (Extended Data Fig. 1). The file was downsampled to 1 variant every 100th using vcftools v.0.1.17<sup>91</sup>, prior to LD estimation. We estimated LD using PLINK v.1.9<sup>83</sup>, and recorded  $R^2$  values >0.05 for pairs up to 15 mb apart or with 10,000 or fewer variants between them. We estimated LD for the unlocalized scaffold 2 of chromosome 13, which contains the morph loci and is ~15 mb in the A assembly. For comparison, we also estimated LD across the first 15 mb of the fully assembled chromosomes (1–12 and X), the main scaffold of chromosome 13, and the unlocalized scaffolds 1, 3 and 4 of chromosome 13.

We used the TE annotations from RepeatModeler v.2.0.1, RepeatMasker v.1.0.93<sup>81</sup> and 'One code to find them all'<sup>94</sup> to quantify TE coverage in chromosome 13 in comparison to the rest of the genome. We divided each chromosome into 1.5 mb windows, and computed the proportion of each window covered by each TE family.

### Evidence of a trans-species polymorphism

We used pool-seq data from the closely related tropical bluetail damselfly (*I. senegalensis*) to determine whether male mimicry has a shared genetic basis in the two species (Extended Data Fig. 1). First, we aligned the short-read data from the two *I. senegalensis* pools (A and O-like) to the A-morph assembly of *I. elegans* using bwa-mem v.0.7.17<sup>78</sup>, and filtered reads with mapping score <20, using Samtools v.1.14<sup>87</sup>. We then quantified read depth as for the *I. elegans* resequencing data (see 'Read-depth analysis'). To confirm that the higher read-depth coverage of the A pool is specific to the putative morph locus, we also plotted the distribution of read-depth differences between O-like and A pools across the rest of the genome and compared it to the morph locus (Supplementary Text 5). Next, we determined if the ~20 kb SV that characterizes A and I females of *I. elegans* is also present in A females of *I. senegalensis*. To do this, we mapped the pool-seq data to the O assembly of *I. elegans* as above, and scanned the region at the start of the scaffold 2 of chromosome 13 for SVs using Samplot v.1.3.0<sup>93</sup>. Finally, we aligned the morph-specific assemblies of *I. senegalensis* to the A assembly of *I. elegans* using nucmer v.4.0.0<sup>75</sup>, and preserving alignments >500 bp with identity >70% (Extended Data Fig. 1). We visualized synteny patterns across the morph locus using the package Rldeogram v.0.2.2<sup>76</sup> in R v.4.2.2<sup>77</sup>.

### Gene content and expression in the morph locus

We assembled transcripts in the A morph genome (Extended Data Fig. 1) to identify potential gene models unique to the A or A and I morphs, which would therefore be absent from the *I. elegans* reference annotation (based on the O haplotype). First, all raw RNA-seq data from *I. elegans* samples were mapped to the A assembly using HISAT2 v.2.2.1<sup>95</sup> and reads with mapping quality <60 were filtered using Samtools v.1.19<sup>87</sup>. Transcripts were then assembled in StringTie v.2.1.4<sup>96</sup> under default options, and merged into a single gtf file. Transcript abundances were quantified using this global set of transcripts as

targets, and a transcript count matrix was produced using the prepDE.py3 script provided with StringTie. Mapped RNA-seq data from *I. senegalensis* were also used to assemble transcripts (Extended Data Fig. 1), but this time the HISAT2 assembly was guided by the annotation based on *I. elegans* data, while allowing the identification of novel transcripts. Transcript abundances were quantified as for *I. elegans*.

We analysed differential gene expression using the package edgeR v.3.36<sup>97</sup> in R v.4.2.2<sup>77</sup>. Transcripts with fewer than one count per million in more than three samples were filtered. Library sizes were normalized across samples using the trimmed mean of M-values method<sup>98</sup>, and empirical Bayes tagwise dispersion<sup>99</sup> was estimated prior to pairwise expression analyses. Differential expression of genes in the morph loci was tested using two-tailed exact tests<sup>100</sup>, assuming negative binomially distributed transcript counts and applying the Benjamini and Hochberg's algorithm to control the false discovery rate<sup>101</sup>.

Nucleotide sequences of all transcripts mapped to the 1.5 mb morph locus in the A assembly were selected. Coding sequences in these transcripts were predicted using Transdecoder v.5.5.0 (<https://github.com/TransDecoder/TransDecoder>). Predicted coding sequences and peptide sequences were read from the assemblies using the genome-based coding region annotation produced with Transdecoder and gffread v.0.12.7<sup>102</sup>. We investigated whether any inferred peptides or transcripts were unique to A or A and I by comparing these sequences to the DTOL reference transcriptome and proteome (based on the O haplotype). We then searched for homologous and annotated proteins in other taxa in the Swissprot database using Blast v.2.9.0<sup>86</sup>. We found three gene models that are protein-coding and present in both A and O females (see 'Results' and Fig. 6). We scanned these protein sequences for functional domains using InterProScan<sup>103</sup> and searched for orthologous groups and functional annotations in eggNOG v.5.0<sup>104</sup>.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Sequencing data from this study have been submitted to the NCBI Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra/>) under BioProject PRJNA940276. For details, please see Supplementary Tables 1 and 2. Morph-specific genome assemblies and intermediate output files required to reproduce the figures in the main text and supporting material are available on Zenodo<sup>105</sup>. Source data are provided with this paper.

### Code availability

All code necessary to reproduce the results of this study can be found on Zenodo at <https://doi.org/10.5281/zenodo.8304055> and Github at <https://github.com/bwillink/Morph-locus>.

### References

- Mank, J. E. Sex chromosomes and the evolution of sexual dimorphism: lessons from the genome. *Am. Nat.* **173**, 141–150 (2009).
- De Lisle, S. P. Understanding the evolution of ecological sex differences: integrating character displacement and the Darwin–Bateman paradigm. *Evol. Lett.* **3**, 434–447 (2019).
- Hopkins, B. R. & Kopp, A. Evolution of sexual development and sexual dimorphism in insects. *Curr. Opin. Genet. Dev.* **69**, 129–139 (2021).
- Jukema, J. & Piersma, T. Permanent female mimics in a lekking shorebird. *Biol. Lett.* **2**, 161–164 (2006).
- Hurtado-Gonzales, J. L. & Uy, J. A. C. Alternative mating strategies may favour the persistence of a genetically based colour polymorphism in a pentamorphic fish. *Anim. Behav.* **77**, 1187–1194 (2009).
- Gosden, T. P. & Svensson, E. I. Density-dependent male mating harassment, female resistance, and male mimicry. *Am. Nat.* **173**, 709–721 (2009).
- Falk, J. J., Rubenstein, D. R., Rico-Guevara, A. & Webster, M. S. Intersexual social dominance mimicry drives female hummingbird polymorphism. *Proc. R. Soc. B* **289**, 20220332 (2022).
- Falk, J. J., Webster, M. S. & Rubenstein, D. R. Male-like ornamentation in female hummingbirds results from social harassment rather than sexual selection. *Curr. Biol.* **31**, 4381–4387 (2021).
- Mank, J. E. Sex-specific morphs: the genetics and evolution of intra-sexual variation. *Nat. Rev. Genet.* **24**, 44–52 (2022).
- Lamichhaney, S. et al. Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). *Nat. Genet.* **48**, 84–88 (2016).
- Andrade, P. et al. Regulatory changes in pterin and carotenoid genes underlie balanced color polymorphisms in the wall lizard. *Proc. Natl Acad. Sci. USA* **116**, 5633–5642 (2019).
- Kim, K.-W. et al. Genetics and evidence for balancing selection of a sex-linked colour polymorphism in a songbird. *Nat. Commun.* **10**, 1852 (2019).
- Woronik, A. et al. A transposable element insertion is associated with an alternative life history strategy. *Nat. Commun.* **10**, 5757 (2019).
- Hendrickx, F. et al. A masculinizing supergene underlies an exaggerated male reproductive morph in a spider. *Nat. Ecol. Evol.* **6**, 195–206 (2022).
- Tuttle, E. M. et al. Divergence and functional degradation of a sex chromosome-like supergene. *Curr. Biol.* **26**, 344–350 (2016).
- Sanchez-Donoso, I. et al. Massive genome inversion drives coexistence of divergent morphs in common quails. *Curr. Biol.* **32**, 462–469 (2022).
- Villoutreix, R. et al. Inversion breakpoints and the evolution of supergenes. *Mol. Ecol.* **30**, 2738–2755 (2021).
- Tunström, K. et al. Evidence for a single, ancient origin of a genus-wide alternative life history strategy. *Sci. Adv.* **9**, eabq3713 (2023).
- Gutiérrez-Valencia, J., Hughes, P. W., Berdan, E. L. & Slotte, T. The genomic architecture and evolutionary fates of supergenes. *Genome Biol. Evol.* **13**, evab057 (2021).
- Sandkam, B. A. et al. Extreme Y chromosome polymorphism corresponds to five male reproductive morphs of a freshwater fish. *Nat. Ecol. Evol.* **5**, 939–948 (2021).
- Kunte, K. et al. Doublesex is a mimicry supergene. *Nature* **507**, 229–232 (2014).
- Feiner, N. et al. A single locus regulates a female-limited color pattern polymorphism in a reptile. *Sci. Adv.* **8**, eabm2387 (2022).
- Neff, B. D. & Svensson, E. I. Polyandry and alternative mating tactics. *Phil. Trans. R. Soc. B* **368**, 20120045 (2013).
- Takahashi, M., Takahashi, Y. & Kawata, M. Candidate genes associated with color morphs of female-limited polymorphisms of the damselfly *Ischnura senegalensis*. *Heredity* **122**, 81–92 (2019).
- Willink, B., Duryea, M. C., Wheat, C. & Svensson, E. I. Changes in gene expression during female reproductive development in a color polymorphic insect. *Evolution* **74**, 1063–1081 (2020).
- Takahashi, M., Okude, G., Futahashi, R., Takahashi, Y. & Kawata, M. The effect of the doublesex gene in body colour masculinization of the damselfly *Ischnura senegalensis*. *Biol. Lett.* **17**, 20200761 (2021).
- Cordero, A., Carbone, S. S. & Utzeri, C. Mating opportunities and mating costs are reduced in androchrome female damselflies, *Ischnura elegans* (Odonata). *Anim. Behav.* **55**, 185–197 (1998).
- Blow, R., Willink, B. & Svensson, E. I. A molecular phylogeny of forktail damselflies (genus *ischnura*) reveals a dynamic macroevolutionary history of female colour polymorphisms. *Mol. Phylogenet. Evol.* **160**, 107134 (2021).

29. Henze, M. J., Lind, O., Wilts, B. D. & Kelber, A. Pterin-pigmented nanospheres create the colours of the polymorphic damselfly *Ischnura elegans*. *J. R. Soc. Interface* **16**, 20180785 (2019).
30. Svensson, E. I., Abbott, J. & Härdling, R. Female polymorphism, frequency dependence, and rapid evolutionary dynamics in natural populations. *Am. Nat.* **165**, 567–576 (2005).
31. Le Rouzic, A., Hansen, T. F., Gosden, T. P. & Svensson, E. I. Evolutionary time-series analysis reveals the signature of frequency-dependent selection on a female mating polymorphism. *Am. Nat.* **185**, E182–E196 (2015).
32. Svensson, E. I., Willink, B., Duryea, M. C. & Lancaster, L. T. Temperature drives pre-reproductive selection and shapes the biogeography of a female polymorphism. *Ecol. Lett.* **23**, 149–159 (2020).
33. Willink, B., Duryea, M. C. & Svensson, E. I. Macroevolutionary origin and adaptive function of a polymorphic female signal involved in sexual conflict. *Am. Nat.* **194**, 707–724 (2019).
34. Price, B. W. et al. The genome sequence of the blue-tailed damselfly, *elegans* (Vander Linden, 1820) [version 1; peer review: awaiting peer review]. *Wellcome Open Res.* **7**, 66 (2022).
35. Voicheck, Y. & Weigel, D. Identifying genetic variants underlying phenotypic variation in plants without complete genomes. *Nat. Genet.* **52**, 534–540 (2020).
36. Parr, M. J. The terminology of female polymorphs of *Ischnura* (Zygoptera: Coenagrionidae). *Int. J. Odonatol.* **2**, 95–99 (1999).
37. Takahashi, Y. & Watanabe, M. Male mate choice based on ontogenetic colour changes of females in the damselfly *Ischnura senegalensis*. *J. Ethol.* **29**, 293–299 (2011).
38. Okude, G. et al. Molecular mechanisms underlying metamorphosis in the most-ancestral winged insect. *Proc. Natl Acad. Sci. USA* **119**, e2114773119 (2022).
39. Laity, J. H., Lee, B. M. & Wright, P. E. Zinc finger proteins: new insights into structural and functional diversity. *Curr. Opin. Struct. Biol.* **11**, 39–46 (2001).
40. Gudmunds, E., Wheat, C. W., Khila, A. & Husby, A. Functional genomic tools for emerging model species. *Trends Ecol. Evol.* **37**, 1104–1115 (2022).
41. Chakraborty, M., Emerson, J. J., Macdonald, S. J. & Long, A. D. Structural variants exhibit widespread allelic heterogeneity and shape variation in complex traits. *Nat. Commun.* **10**, 4872 (2019).
42. Quan, C., Lu, H., Lu, Y. & Zhou, G. Population-scale genotyping of structural variation in the era of long-read sequencing. *Comput. Struct. Biotechnol. J.* **37**, 1104–1115 (2022).
43. Schrader, L. & Schmitz, J. The impact of transposable elements in adaptive evolution. *Mol. Ecol.* **28**, 1537–1549 (2019).
44. Catlin, N. S. & Josephs, E. B. The important contribution of transposable elements to phenotypic variation and evolution. *Curr. Opin. Plant Biol.* **65**, 102140 (2022).
45. Zhang, J. et al. Autotetraploid rice methylome analysis reveals methylation variation of transposable elements and their effects on gene expression. *Proc. Natl Acad. Sci. USA* **112**, E7022–E7029 (2015).
46. Diehl, A. G., Ouyang, N. & Boyle, A. P. Transposable elements contribute to cell and species-specific chromatin looping and gene regulation in mammalian genomes. *Nat. Commun.* **11**, 1796 (2020).
47. Sundaram, V. & Wysocka, J. Transposable elements as a potent source of diverse cis-regulatory sequences in mammalian genomes. *Phil. Trans. R. Soc. B* **375**, 20190347 (2020).
48. Jangam, D., Feschotte, C. & Betrán, E. Transposable element domestication as an adaptation to evolutionary conflicts. *Trends Genet.* **33**, 817–831 (2017).
49. McCue, A. D. & Slotkin, R. K. Transposable element small RNAs as regulators of gene expression. *Trends Genet.* **28**, 616–623 (2012).
50. Kiuchi, T. et al. A single female-specific piRNA is the primary determinant of sex in the silkworm. *Nature* **509**, 633–636 (2014).
51. Van Gossum, H., Stoks, R. & De Bruyn, L. Frequency-dependent male mate harassment and intra-specific variation in its avoidance by females of the damselfly *Ischnura elegans*. *Behav. Ecol. Sociobiol.* **51**, 69–75 (2001).
52. Takahashi, Y., Kagawa, K., Svensson, E. I. & Kawata, M. Evolution of increased phenotypic diversity enhances population performance by reducing sexual harassment in damselflies. *Nat. Commun.* **5**, 4468 (2014).
53. Rowe, L., Chenoweth, S. F. & Agrawal, A. F. The genomics of sexual conflict. *Am. Nat.* **192**, 274–286 (2018).
54. Sayadi, A. et al. The genomic footprint of sexual conflict. *Nat. Ecol. Evol.* **3**, 1725–1730 (2019).
55. Wellenreuther, M. et al. Molecular and ecological signatures of an expanding hybrid zone. *Ecol. Evol.* **8**, 4793–4806 (2018).
56. Okude, G., Fukatsu, T. & Futahashi, R. Interspecific crossing between blue-tailed damselflies *Ischnura elegans* and *I. senegalensis* in the laboratory. *Entomol. Sci.* **23**, 165–172 (2020).
57. Montgomery, E. A., Huang, S. M., Langley, C. H. & Judd, B. H. Chromosome rearrangement by ectopic recombination in *Drosophila melanogaster*: genome structure and evolution. *Genetics* **129**, 1085–1098 (1991).
58. Delprat, A., Negre, B., Puig, M. & Ruiz, A. The transposon *Galileo* generates natural chromosomal inversions in *Drosophila* by ectopic recombination. *PLoS ONE* **4**, e7883 (2009).
59. Andrés, J. A. & Cordero, A. The inheritance of female colour morphs in the damselfly *Ceragrion tenellum* (Odonata, Coenagrionidae). *Heredity* **82**, 328–335 (1999).
60. Kent, T. V., Uzunović, J. & Wright, S. I. Coevolution between transposable elements and recombination. *Phil. Trans. R. Soc. B* **372**, 20160458 (2017).
61. Thompson, M. J. & Jiggins, C. Supergenes and their role in evolution. *Heredity* **113**, 1–8 (2014).
62. Willink, B. & Svensson, E. I. Intra- and intersexual differences in parasite resistance and female fitness tolerance in a polymorphic insect. *Proc. R. Soc. B* **284**, 20162407 (2017).
63. Abbott, J. K. & Gosden, T. P. Correlated morphological and colour differences among females of the damselfly *Ischnura elegans*. *Ecol. Entomol.* **34**, 378–386 (2009).
64. West-Eberhard, M. J. *Developmental Plasticity and Evolution* (Oxford Univ. Press, 2003).
65. Svensson, E. I., Abbott, J. K., Gosden, T. P. & Coreau, A. Female polymorphisms, sexual conflict and limits to speciation processes in animals. *Evol. Ecol.* **23**, 93–108 (2009).
66. Colgan, D. J. et al. Histone H3 and U2 snRNA DNA sequences and arthropod molecular evolution. *Aust. J. Zool.* **46**, 419–437 (1998).
67. Shafin, K. et al. Nanopore sequencing and the Shasta toolkit enable efficient de novo assembly of eleven human genomes. *Nat. Biotechnol.* **38**, 1044–1053 (2020).
68. Phillippy, A. M., Schatz, M. C. & Pop, M. Genome assembly forensics: finding the elusive mis-assembly. *Genome Biol.* **9**, R55 (2008).
69. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
70. Shafin, K. et al. Haplotype-aware variant calling with PEPPER-Margin-DeepVariant enables high accuracy in nanopore long-reads. *Nat. Methods* **18**, 1322–1332 (2021).
71. Guan, D. et al. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36**, 2896–2898 (2020).
72. Zimin, A. V. et al. Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. *Genome Res.* **27**, 787–792 (2017).
73. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).

74. Alonge, M. et al. Automated assembly scaffolding elevates a new tomato system for high-throughput genome editing. *Genome Biol.* **23**, 258 (2022).
75. Delcher, A. L., Phillippy, A., Carlton, J. & Salzberg, S. L. Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res.* **30**, 2478–2483 (2002).
76. Hao, Z. et al. Rideogram: drawing SVG graphics to visualize and map genome-wide data on idiograms. *PeerJ Comput. Sci.* **6**, e251 (2020).
77. R Core Team R: *A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2021).
78. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at <https://arxiv.org/abs/1303.3997> (2013).
79. Van der Auwera, G. A. et al. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinform.* **43**, 11.10.1–11.10.33 (2013).
80. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
81. Smith, A. F. A., Hubley, R. & Green, P. RepeatMasker Open-4.0 <http://www.repeatmasker.org> (2013–2015).
82. Girgis, H. Z. Red: an intelligent, rapid, accurate tool for detecting repeats de-novo on the genomic scale. *BMC Bioinform.* **16**, 227 (2015).
83. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
84. Kokot, M., Dlugosz, M. & Deorowicz, S. KMC 3: counting and manipulating *k*-mer statistics. *Bioinformatics* **33**, 2759–2761 (2017).
85. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824 (2012).
86. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinform.* **10**, 421 (2009).
87. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
88. Pedersen, B. S. & Quinlan, A. R. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* **34**, 867–868 (2018).
89. Korunes, K. L. & Samuk, K. pixy: unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Mol. Ecol. Resour.* **21**, 1359–1368 (2021).
90. Hudson, R. R., Slatkin, M. & Maddison, W. P. Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**, 583–589 (1992).
91. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
92. Sedlazeck, F. J. et al. Accurate detection of complex structural variations using single-molecule sequencing. *Nat. Methods* **15**, 461–468 (2018).
93. Belyeu, J. R. et al. Samplot: a platform for structural variant visual validation and automated filtering. *Genome Biol.* **22**, 161 (2021).
94. Bailly-Bechet, M., Haudry, A. & Lerat, E. ‘One code to find them all’: a perl tool to conveniently parse RepeatMasker output files. *Mob. DNA* **5**, 13 (2014).
95. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
96. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
97. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. EdgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
98. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).
99. Robinson, M. D. & Smyth, G. K. Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics* **23**, 2881–2887 (2007).
100. Robinson, M. D. & Smyth, G. K. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics* **9**, 321–332 (2008).
101. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
102. Pertea, G. & Pertea, M. GFF utilities: GffRead and GffCompare. *F1000Research* **9**, 304 (2020).
103. Jones, P. et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
104. Huerta-Cepas, J. et al. eggNOG5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309–D314 (2019).
105. Willink, B. et al. Data from: The genomics and evolution of inter-sexual mimicry and female-limited polymorphisms in damselflies. *Zenodo* <https://zenodo.org/records/8304153> (2023).
106. Takahashi, Y. Testing negative frequency-dependent selection: linking behavioral plasticity and evolutionary dynamics. *Bull. Kanto Branch Ecol. Soc. Jpn.* **59**, 8–14 (2011).

## Acknowledgements

We thank H. Dort, R. A. Steward, J. Haushofer, P. de Sessions and the Monteiro Lab for suggestions and helpful discussions. We are also grateful to M. P. Celorio-Mancera and H. M. Low for invaluable technical support. A. Monteiro hosted B.W. at the National University of Singapore during part of the duration of this study. Computation and data handling were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) through the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX), under projects 2022/6-230 and 2022/5-419 awarded to E.I.S. Specimens of *I. senegalensis* were collected in Singapore, under a research permit (NP/RP22-015b) from the National Parks Board, Singapore. This work was supported by an International Postdoc Grant from the Swedish Research Council (VR) (grant no. 2019-06444 to B.W.). Funding was also provided by the Swedish Research Council (VR) (grant no. 2017-04386 to C.W.W. and grant no. 2016-03356 to E.I.S.), by the Stina Werners Foundation (grant no. 2018-017 to E.I.S.) and Erik Philip Sörensens Stiftelse (grant no. 2019-033 to E.I.S.). S.N. was funded by a scholarship grant for Master's students from Sven and Lily Lawski's Foundation.

## Author contributions

B.W. conceived the study with input from C.W.W. E.I.S. organized field work in the long-term population study of *I. elegans* during 2019 and 2020, and planned and prepared the outdoor rearing experiments. E.I.S. and S.N. collected DNA and RNA samples of *I. elegans*. M.T. and Y.T. collected samples for pool sequencing of *I. senegalensis*, and B.W. collected samples of *I. senegalensis* in Singapore. S.N., B.W. and K.T. conducted laboratory work on *I. elegans*. M.T., Y.T. and B.W. conducted laboratory work on *I. senegalensis*. B.W. analysed the data with input from C.W.W., K.T., R.C. and T.L. B.W. wrote the paper with contributions from all authors.

## Funding

Open access funding provided by Stockholm University.

## Competing interests

The authors declare no competing interests.

## Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41559-023-02243-1>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41559-023-02243-1>.

**Correspondence and requests for materials** should be addressed to Beatriz Willink.

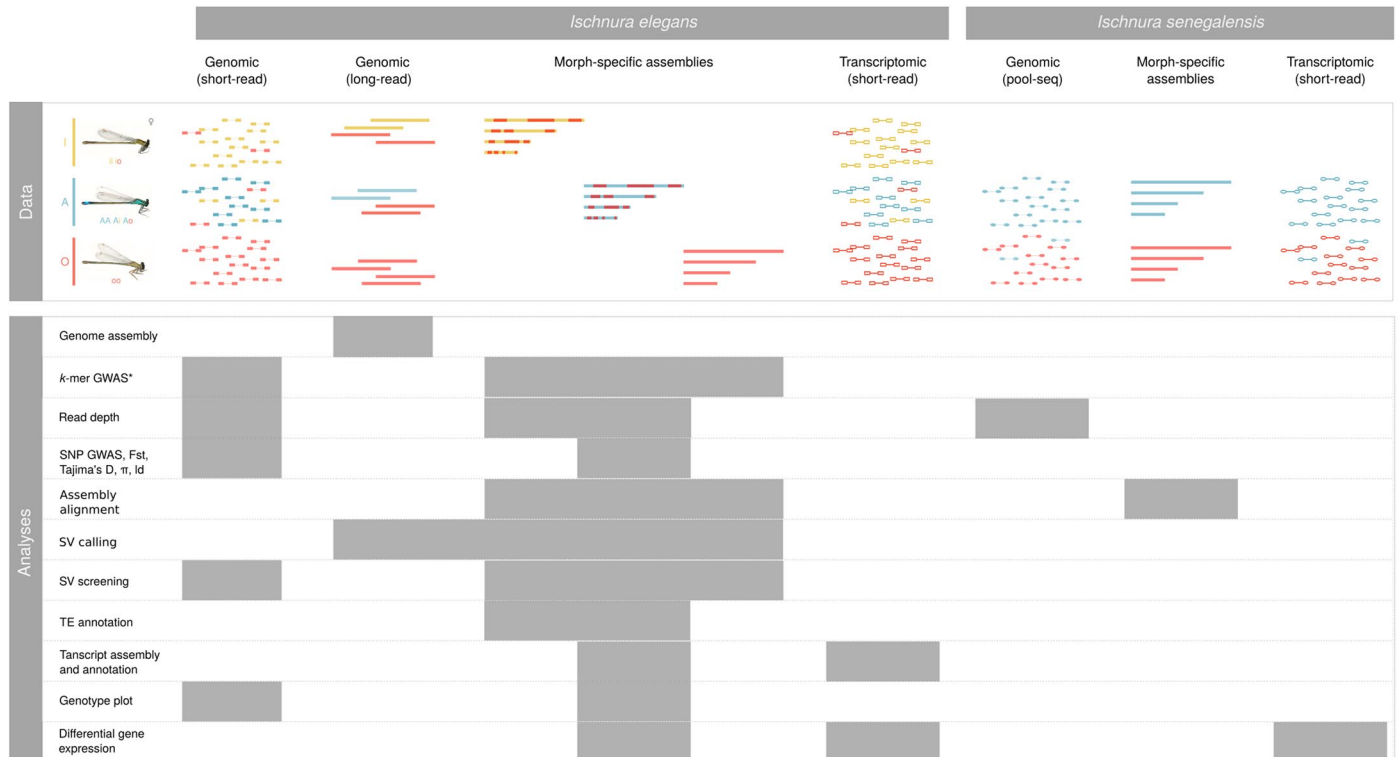
**Peer review information** *Nature Ecology & Evolution* thanks Riddhi Deshmukh, Elina Immonen and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

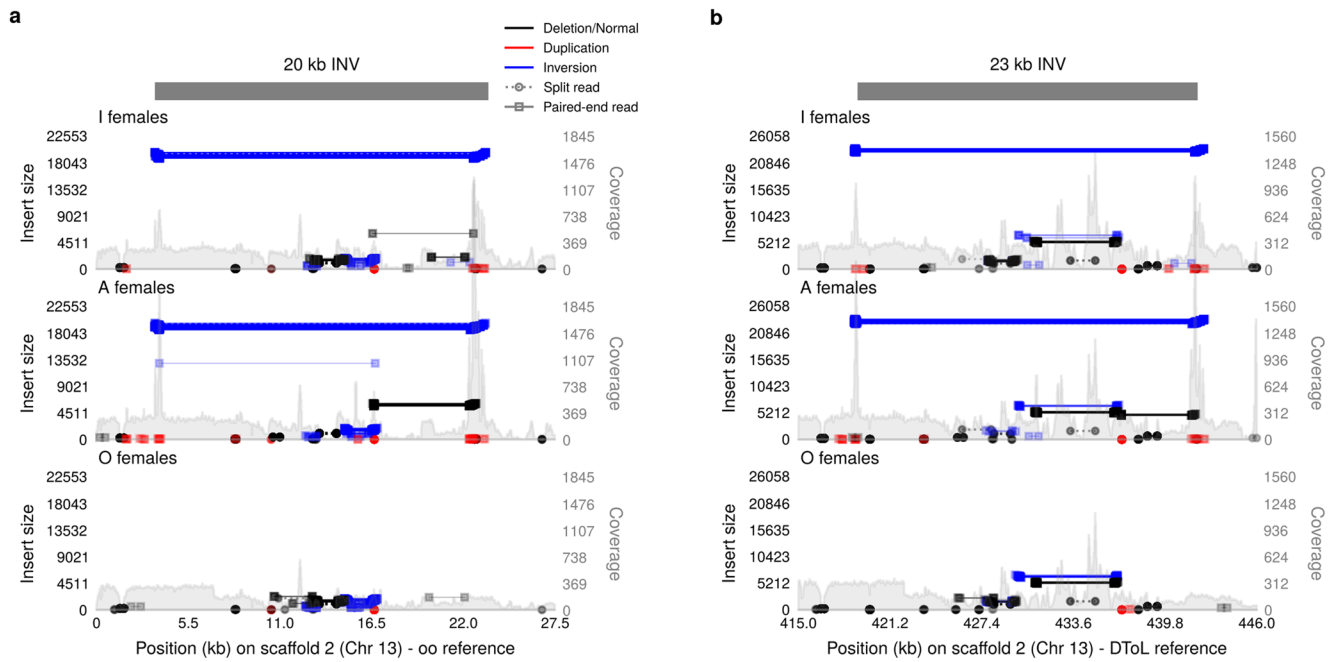
© The Author(s) 2023



**Extended Data Fig. 1 | Outline of data and analyses used in this study.** For our main study species *Ischnura elegans*, we obtained short-read genomic data from 19 field-caught females per morph, and long-read genomic data from three females with genotypes *Ao*, *Io*, and *oo*. The long-read samples were used to assemble morph-specific genomes, scaffolded against the Darwin Tree of Life reference assembly. We obtained whole-thorax RNAseq data from females of each morph in both sexually immature and sexually mature colour phases ( $n = 3$  of each morph and colour phase). Immature and mature males ( $n = 3$  of each) were also sampled for whole-thorax RNAseq data. We used short-read pool-seq data ( $n = 30$  individuals of each morph per pool) of the close relative *Ischnura*

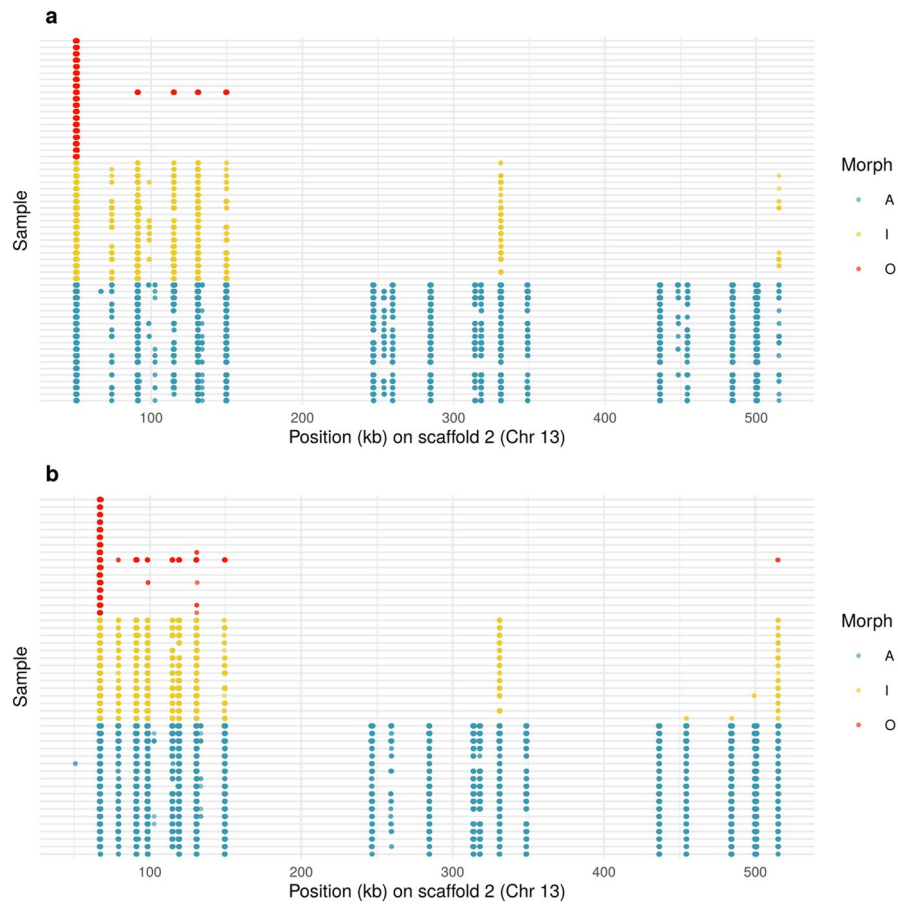
*senegalensis* to investigate whether the female polymorphisms in both species share a genomic basis. We also analysed expression levels of candidate genes in this species, using samples from a previously published study<sup>38</sup>, which produced transcriptomic data from four body parts (head, thorax, wing and abdomen) of each A females, O females and males ( $n = 2$ ), sampled at adult emergence and two days thereafter. The *k*-mer based GWAS is reference-free, but significant *k*-mers were mapped to the morph-specific assemblies to determine their chromosomal context. Damselfly images adapted from ref. 25 under a Creative Commons licence CC BY 4.0.



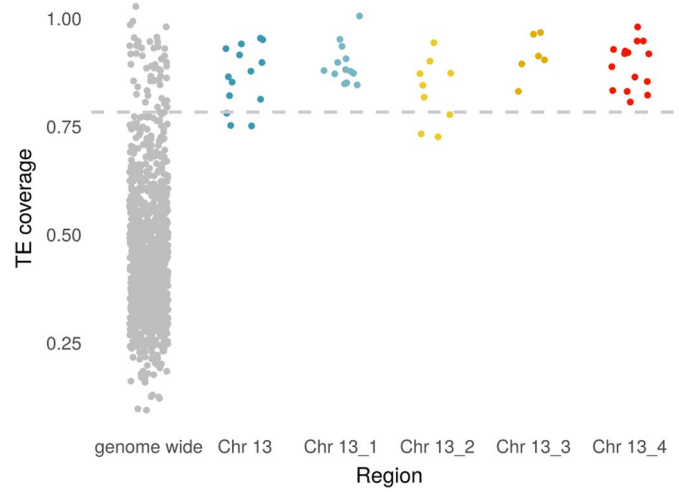


**Extended Data Fig. 2 | An inversion signature differentiates A and I individuals from the O morph.** Read mapping and sample coverage at the start of the scaffold 2 of chromosome 13 in **a** our O assembly and **b** the DTOL reference assembly, showing a signature of a ~20 kb inversion in A and I

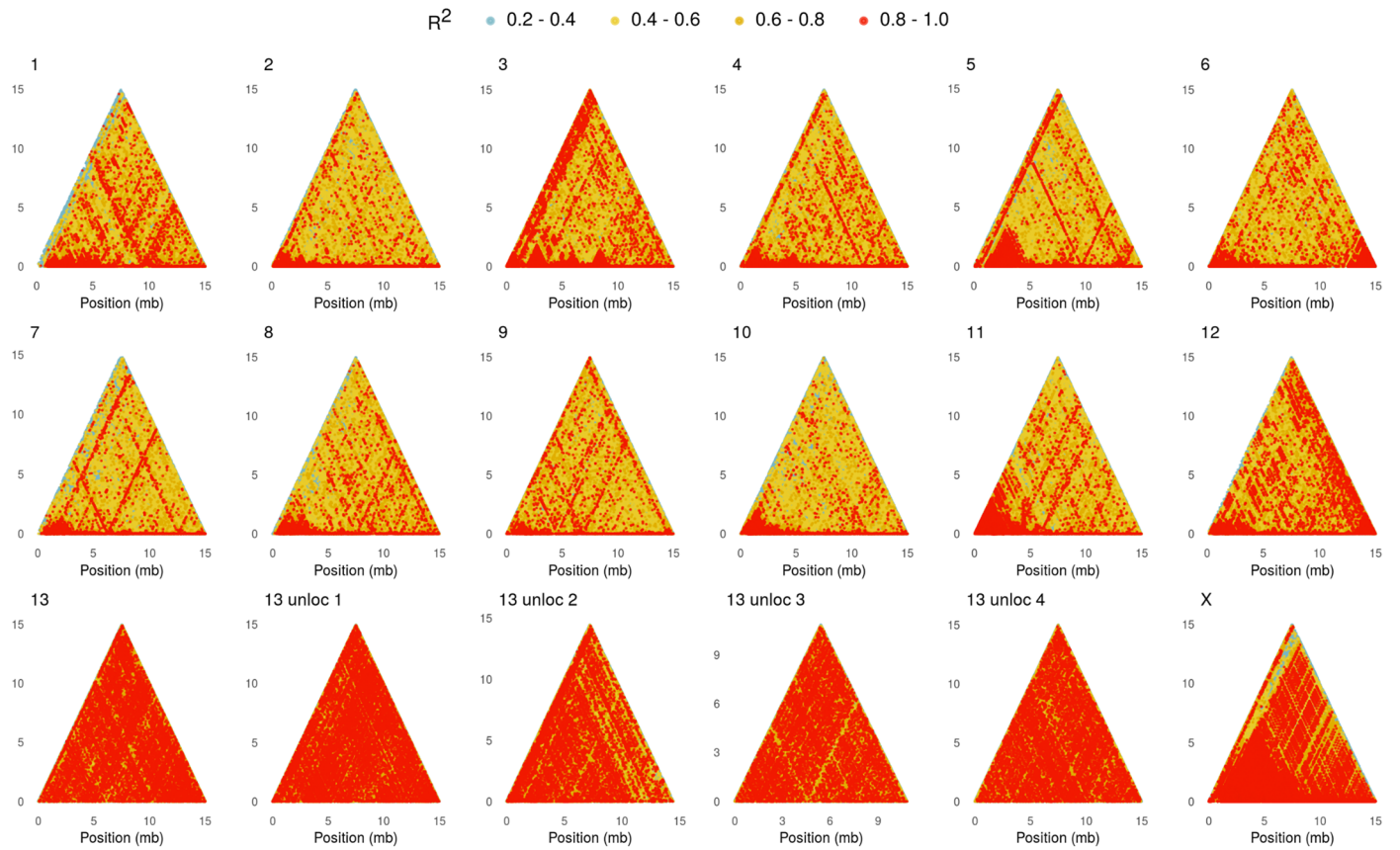
I samples. A single O sample also exhibited this signature but was excluded here for clarity (see Supporting Text 3). Note that the first 415 kb of the reference DTOL assembly are absent in our scaffolded O assembly, and therefore the x-axis is shifted by 415 kb in **b**.



**Extended Data Fig. 3 | The A and I reads mapped to inversion break points on the O assembly (see Extended Data Fig. 2) map to multiple locations on the A assembly. **a** Reads from the first inversion breakpoint. **b** Reads from the second inversion breakpoint. Each row represents a sample and each circle an individual read. The x-axis corresponds to coordinates on the A assembly.**

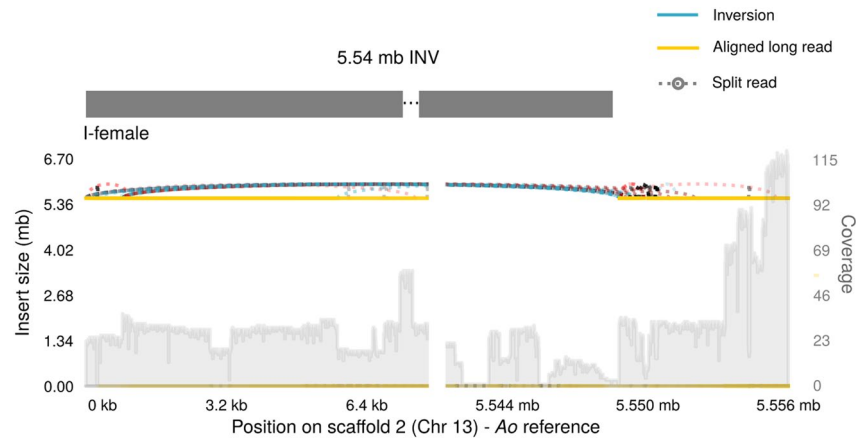


**Extended Data Fig. 4 | Proportion of TE content in non-overlapping 1.5 mb regions.** The gray dots correspond to genomic windows outside chromosome 13. The main assembly and the unlocalized scaffolds of chromosome 13 are depicted with different colours. The dashed line marks the 95 percentile of TE coverage across all windows.



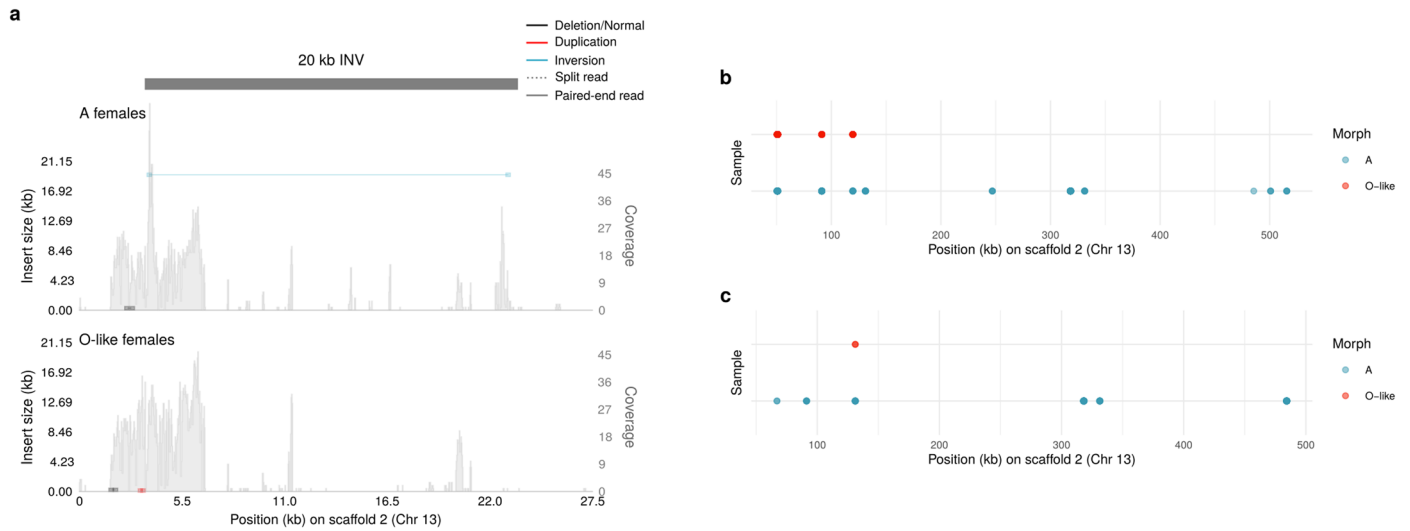
**Extended Data Fig. 5 | Linkage disequilibrium (LD) in the genome of *Ischnura elegans*.** LD estimates are shown for the first 15 mb of each chromosome and all unlocalized scaffolds of chromosome 13. The morph locus is found in the first

~1.5 mb of the unlocalized scaffold 2 of chromosome 13, which has a total size of ~15 mb. Each dot represent the square correlation coefficient ( $R^2$ ) between two variant sites on the x axis, separated by the number of sites indicated in the y axis.



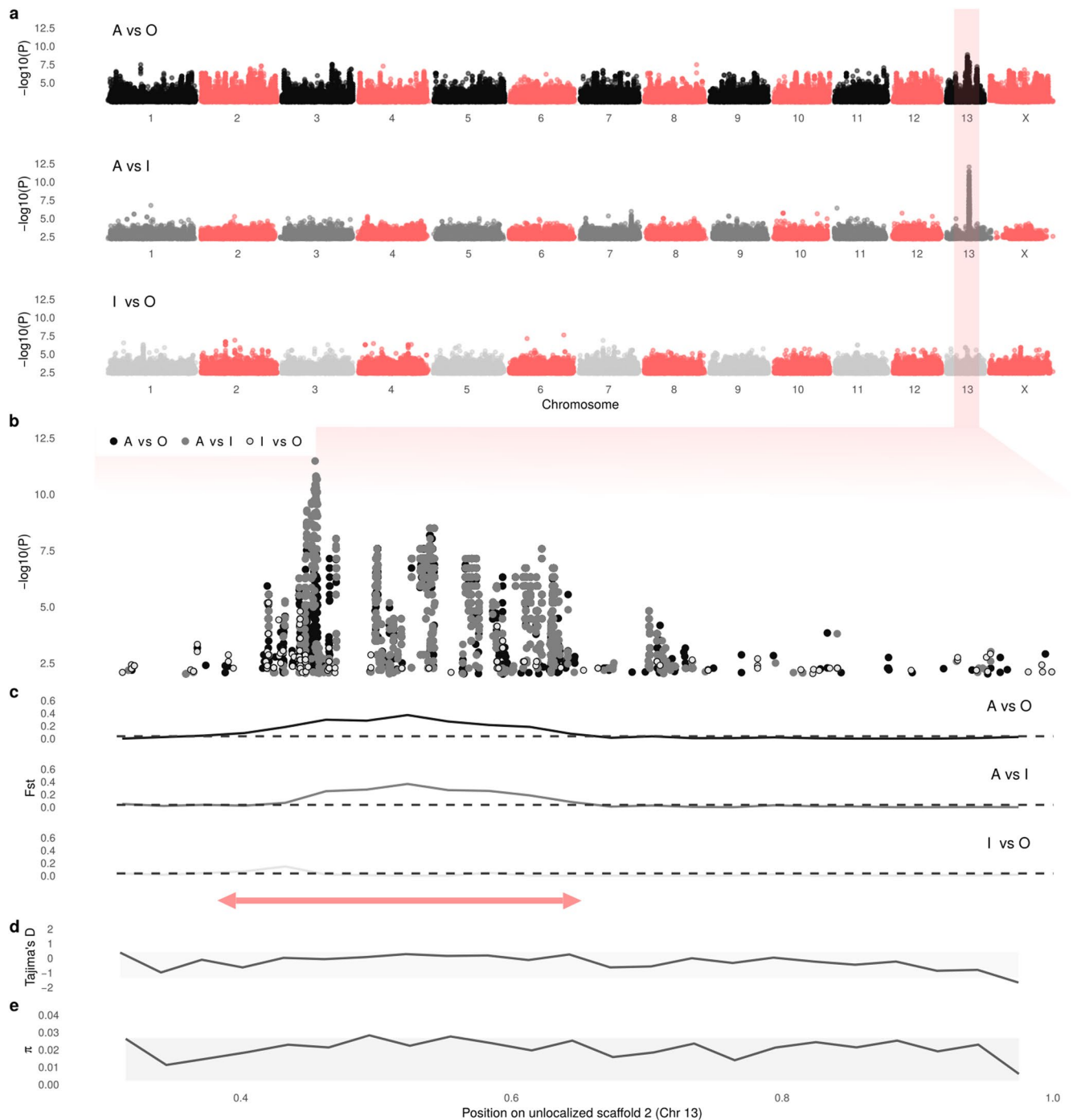
**Extended Data Fig. 6 | Evidence of a translocation between A and I haplotypes.** Mapping and coverage of long reads from an *I*o sample across the first 5.6 mb of the unlocalized scaffold 2 of chromosome 13 in the A assembly,

showing a signature consistent with either a 5.54 mb inversion or a translocation of inverted A content. Absence of morph divergence beyond -1.5 mb on the A assembly supports the translocation scenario.



**Extended Data Fig. 7 | Structural variants between A and O-like females of *I. senegalensis* along the morph locus identified in *I. elegans*.** **a** Read mapping and sample coverage of *I. senegalensis* pool-seq data at the start of the unlocalized scaffold 2 of chromosome 13 in the O assembly of *I. elegans*. The same ~20 kb inversion signature is found in A and I samples of *I. elegans*

(see Extended Data Fig. 2). **b-c** The A-pool reads mapped to break points on the O assembly map to multiple locations on the A assembly. **b** Reads from the first breakpoint. **c** Reads from the second breakpoint. Each row represents a pool of *I. senegalensis* and each circle an individual read. The x-axis corresponds to the A assembly of *I. elegans*.



### Extended Data Fig. 8 | Morph divergence using the DTOL assembly

**(O haplotype) as mapping reference.** **a** SNP-based genome-wide associations in all pairwise analyses between morphs. Genomic DNA from 19 wild-caught females of each colour morph and of unknown genotype was extracted and sequenced for these analyses. Illumina short reads were aligned against the DTOL reference assembly. **b** A closer look of the SNP associations on the unlocalized scaffold 2 of chromosome 13, which contained all highly significant SNPs. The y axis in **a** and **b** indicates unadjusted  $-\log_{10}$  P-values calculated from chi-squared tests.

**c**  $F_{st}$  values averaged across 30 kb windows for the same pairwise comparisons as in the SNP based GWAS. The dashed line marks the 95 percentile of all non-zero  $F_{st}$  values across the entire genome. The red double arrow shows the region of elevated divergence between A and both O and I samples. Population-level estimates of **d** Tajima's D, and **e** nucleotide diversity ( $\pi$ ) averaged across 30 kb windows. The shaded area contains the 5–95 percentile of all genome-wide estimates.

**Extended Data Table 1 | Significant *k*-mers associated with morph comparisons in *I. elegans***

<b>k-mer</b>	<b>A vs O</b>	<b>A vs I</b>	<b>I vs O</b>
Total number	568,039	508,031	85,134
A assembly	435,509	383,679	45,580
A Chr 13_2	427,606	376,075	44,990
I assembly	46,733	1,111	49,093
I Chr 13_2	45,819	375	48,484
O assembly	1,478	762	1,452
O Chr 13_2	276	92	72
DToL primary assembly	915	756	676
DToL Chr 13_2	134	115	14
DToL haplotigs	542,571	489,912	66,861
DToL RAPID 106	539,588	486,943	65,999

For each comparison (A vs O, A vs I and A and I vs O), we show the total number of significant *k*-mers, and the total number of significant *k*-mers that map without any mismatching position to morph-specific reference assemblies. Of the mapping *k*-mers, we then show the number located in the unlocalized scaffold 2 of chromosome 13, which includes the putative morph locus. For the DToL assembly, we show the number of significant *k*-mers mapping to both the primary assembly (capturing the O allele) and the haplotigs, where the haplotig RAPID\_106 comprises the A allele (see Supporting Text 2).



## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a | Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

Software used for base calling in raw ONT data: Guppy v 4.0.11, Guppy v 5.0.11, Guppy v 6.1.5. More details are available in the Methods section.

Data analysis

All code used for data analysis in this study is available at github: <https://github.com/bwillink/Morph-locus>

Software used for genome assembly and polishing: Shasta v 0.7.0, minimap2 v 2.22-r1110 83, PEPPER-Margin-DeepVariant pipeline r0.4, purge\_dups v 0.0.3, MaSuRCA v 4.0.4 86, BUSCO v 5.0.0, RagTag v 2.10

Software used for genome alignment: nucmer v 4.0.0

Software used for genomic read mapping and pre-processing: bwa-mem v 0.7.17, minimap2 v 2.22-r1110, samtools v 1.9, samtools v 1.14, Picard tools v 2.23.4, GATK v 4.2.0.0

Software used for variant calling and filtering: bcftools v 1.12, RepeatMasker v 1.0 93, Red v 0.0.1

Software used for GWAS: PLINK v 1.9

Software used for k-mer based GWAS: KMC v 3.1.0, GEMMA v 0.98.5, Blast v 2.2.28

Software used for computing read-depth coverage: mosdepth v 0.2.8

Software used for population genetics analysis: pixy v 1.2.5 (Fst and pi), vcftools v 0.1.17 (Tajima's D)

Software used for calling structural variants: Sniffles v 1.0.10, SamPlot v 1.3.0

Software used for estimating linkage disequilibrium: PLINK v 1.9

Software used for TE annotation: RepeatModeler v 2.0.1, RepeatMasker v 1.0.93, "One code to find them all" v 1.0

Software used for RNAseq read mapping, transcript assembly and count: HISAT2 v 2.2.1, samtools v 1.9, StringTie v 2.1.4

Software used for expression analysis: R v 4.2.2, edgeR v 3.36

Software used for transcript annotation: Transdecoder v 5.5.0, gffread v 0.12.7, Blast v 2.9.0, OrthoFinder v 2.5.2

Software used for PCA: PLINK v 1.9

Software used for plotting: SamPlot v 1.3.0, R v 4.2.2, R packages: RIdeogram v 0.2.2, ggplot2 v 3.4.1, wesanderson v 0.3.6, tidyverse v 1.3.1, dplyr v 1.1.0, gridExtra v 2.3, ggtree v 3.6.2, GenotypePlot v 0.2.1, vcfR v 1.12.0, viridis v 0.6.2, Gviz v 1.38.4

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Sequencing data from this study have been submitted to the NCBI Sequence Read Archive (SRA) (<https://www.ncbi.nlm.nih.gov/sra/>) under accession number PRJNA940276. For individual sample accessions please see Supporting Text 1 and Tables S1 and S2 in the Supplementary Material.

## Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

Population characteristics

Recruitment

Ethics oversight

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences     Behavioural & social sciences     Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

This study consists of several parts with different experimental units. For an overview of analyses and data types see Extended Data Figure 1.

1) De novo genome assembly for female morphs of *Ischnura elegans* and *Ischnura senegalensis*. Assemblies were based on one individual sample per morph per species.

2) GWAS, read-depth analysis, population genetics analysis, linkage disequilibrium and SV analysis in *Ischnura elegans*. These analyses were based on 19 whole-genome resequencing samples of each female morph. See Table S1 for more information on each

	sample. Read-depth analyses and SV calling also used the three long-read sequencing samples used for genome assembly in 1).
	3) Read-depth and SV analysis in <i>Ischnura senegalensis</i> . These analyses were based on two pool-seq DNA samples. Each pool consisted of 30 females of each morph of <i>Ischnura senegalensis</i> .
	4) Expression analysis in <i>Ischnura elegans</i> and <i>Ischnura senegalensis</i> . The analysis in <i>Ischnura elegans</i> was based on 24 RNAseq individual samples, including six A-females, six I-females, six O-females and six males. For each sex/morph category with six samples, three were sexually immature adults and three were sexually mature adults. This information is also available on Table S2. We also show publicly available data from a previous study on <i>Ischnura senegalensis</i> (Okude et al. PNAS 119, e2114773119). These data correspond to six individuals, two males, two A-females, and two O-females. For each sex/morph category with two individuals, one was sampled upon adult emergence and the other one was sampled two days after emergence. For each of these individuals the head, thorax, abdomen and wings were collected and sequenced separately.
Research sample	Samples were all adult damselflies collected in the field in Southern Sweden ( <i>I. elegans</i> ), Japan ( <i>I. senegalensis</i> , pool-seq data) and Singapore ( <i>I. senegalensis</i> , long-read data), as specified in the Methods section. Samples were all female, except for the males used for expression analysis as explained in the "Study description".
Sampling strategy	Individuals were collected upon encounter in the field. There was no statistical analysis to predetermine sample size. The choice of sample size was based on available funding and previous studies with similar types of data.
Data collection	EIS and SN collected all <i>I. elegans</i> samples and recorded their sex, age and morph based on visual inspection. YT and MT collected the pool-seq samples of <i>I. senegalensis</i> and recorded the sex and morph of individuals to be pooled based on visual inspection. BW collected the samples for de novo genome assembly in <i>I. senegalensis</i> and recorded the sex and morph of individuals based on visual inspection.
Timing and spatial scale	<i>I. elegans</i> samples were collected in the summers (June-July) of 2019 (resequencing and RNAseq), and 2020 (genome assembly). <i>I. senegalensis</i> samples were collected in May 2016 (pool-seq) and May and September 2022 (genome assembly). Exact sampling dates for all data except the pool-seq data are available through the Biosample accessions provided in Table S1, Table S2 and the Methods section. Given the goals of this study, the frequency of sampling is not expected to impact any of our results.
Data exclusions	No samples were excluded. SNP data was filtered based on quality and repetitive-content annotations, as explained in the Methods section.
Reproducibility	This study contains no experimental findings.
Randomization	Randomization was not applicable in this study.
Blinding	Blinding was not applicable in this study.
Did the study involve field work?	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No

## Field work, collection and transport

Field conditions	Fieldwork consisted only of specimen collection. It was conducted in rural and urban ponds and streams during morning hours on sunny days.
Location	Information on the location of origin of the samples is available through the Biosample accessions on NCBI. See Table S1, Table S2 and the Methods section.
Access & import/export	Collecting permits for insect specimens are only required in Singapore, where research was conducted under a permit from the National Parks Board (NP/RP22-015b). Sequencing was conducted in all cases in the country of origin of the samples, so there was no need for import/export permits. Field sites consisted of ponds and streams in urban or agricultural landscapes. The sites can be easily accessed by researchers and the general public.
Disturbance	The only disturbance caused by the study could have been damage to the surrounding vegetation of ponds and streams as researchers walked and collected specimens. This disturbance is minimal.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials &amp; experimental systems

n/a	Involvement
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

## Methods

n/a	Involvement
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals	This study did not involve laboratory animals
Wild animals	Damselfly specimens were collected in the field for sequencing. Individuals were euthanized by freezing or after being submerged in 95% ethanol or liquid nitrogen. Some individuals were euthanized in the field (for RNAseq analysis) while the others were brought to laboratory and immediately preserved.
Reporting on sex	This study addresses the genomics and evolution of a female-limited colour polymorphism, so most individual samples are female. Where data was collected for males also (for RNAseq analysis) the sex of each sample is indicated (See Table S2).
Field-collected samples	Field collected specimens were euthanized in the field or immediately after being brought to a laboratory. Specimens were immediately stored at -20 C or -80 C.
Ethics oversight	No ethical approval or guidance are required for work with insect samples in Sweden, Japan or Singapore.

Note that full information on the approval of the study protocol must also be provided in the manuscript.