High levels of chromosomal synteny in 250 million year old groups of dragonflies and damselflies (Insecta:Odonata)

Ethan Tolman (etolman@amnh.org)
American Museum of Natural History  https://orcid.org/0000-0002-2594-2833

Christopher Beatty
Stanford University

Jonas Bush
Brigham Young University

Manpreet Kohli
https://orcid.org/0000-0002-3896-4513

Paul Frandsen
Brigham Young University  https://orcid.org/0000-0002-4801-7579

J. Gosnell
Baruch College

Jessica Ware
https://orcid.org/0000-0002-4066-7681

Article

Keywords:

Posted Date: March 21st, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2621772/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

Using recently published chromosome-length genome assemblies of damselfly species *Ischnura elegans* and dragonfly species *Pantala flavescens* and *Tanypteryx hagenii*, we demonstrate that the autosomes of Odonata show a high level of conservation, despite 250 million years of separation. In the four genomes discussed here, our results show that all autosomes have a clear homolog to the ancestral karyotype. Despite clear synteny, we demonstrate that different factors, including concentration of repeat dynamics, GC content, and the relative proportion of coding sequence all influence the amount of syntenic conservation across chromosomes, and that the influence of these factors differ among species. Micro- and sex chromosomes in Odonata do not share the same level of synteny as autosomes. Of the four species sampled, the genome of the Black Petaltail, which diverged from its sister species 70 million years ago, is a clear outlier, showing similarities to other long-lived lineages.

Introduction

Chromosome evolution patterns vary drastically across the tree of life and many studies have explored this in Eumetaza. Most of this work has focused on vertebrates, such as mammals, amphibians, sharks, or birds. In mammals, the rate of chromosome evolution is faster than in most other eumetazoan lineages. Even within mammalia the rate of chromosome evolution is fastest in genera that either form clans or harems, display limited adult vagility, are strongly territorial, or have a patchy geographic distribution. Chromosomal architecture can vary dramatically within and among lineages, for example, primates have several highly conserved chromosomes, while other chromosomes have undergone numerous inversion events. In amphibians, genome duplication and spontaneous polyploidy are more frequent than in mammals, and sex chromosomes are especially fluid. It has been recently discovered that at least one elasmobranch, the white-spotted bamboo shark, has experienced a number of rearrangements, with 13 of the 51 chromosomes evolving more quickly than the others. Chromosome evolution varies dramatically across the ray-finned fishes, class Actinopterygii. Some species have maintained the ancestral karyotype for over 300 million years, while others have undergone such a high number of translocations and other rearrangements that homology with the ancestral karyotype can no longer be identified. As a general rule, it is possible that the karyotype of fresh water fishes evolves at a faster rate than their saltwater counterparts. Amongst vertebrates, birds are an outlier in that their chromosomes reflect very slow evolutionary rates, with few chromosomal rearrangements. First radiating in the late Cretaceous (over 65 mya), birds show a much higher level of chromosome conservation than other vertebrate taxa.

Invertebrate taxa have been less well-studied in terms of chromosome evolution than vertebrates, including Insecta, the most diverse group of animals on the planet. This knowledge gap is problematic considering the rich global biodiversity and long evolutionary history (with flying insects having radiated ~ 400 mya) of insects. Existing insect karyotype research suggests that trends in insect chromosome evolution vary greatly across orders. For example, in the Paraneoptera, a group which comprises Hemiptera ("true bugs") and Psocidea (lice), there have been many transitions resulting in both the gain and loss of holokinetic chromosomes (chromosomes which lack a centromere). This is not seen in the other Paraneopteran order, Thysanoptera (thrips). Across the ~ 28 orders of insects chromosome number varies greatly, but, of the groups studied, the most diverse karyotypes occur in Aphidoidea (aphids), a lineage within the Hemiptera, where chromosome number ranges from 2n = 4 to 2n = 192. Several studies have examined holometabolous insect (those which undergo complete metamorphosis) karyotypes and, in the case of scorpionflies, or Mecoptera, the family Panorpidae have haploid chromosomes which range from n = 17 to n = 24 and lineage differentiation in this family is closely related to chromosome evolution. Throughout the 140 million-year history of Lepidoptera (butterflies and moths), most species have nearly identical haploid chromosome counts, although a handful of species vary drastically from the ancestral karyotype due to numerous fusion and fission events, with translocation likely playing a very limited role in the evolution of Lepidopteran chromosomes.

To better infer the evolution of pterygota (winged insects) chromosome sampling should be done across the lass Insecta, including taxa that are non-Holometabola and non-Paraneoptera. The Odonata (dragonflies and damselflies) are one of the earliest diverging winged insect lineages (~325my). The ancestor to these insects was likely the first lineage to have developed flight, an energetically costly key innovation that is considered to be a main driver of evolution in Pterygota. As such, Odonata are a key model for understanding the earliest evolution of flight. Odonate karyotypes have been studied for over a century. The chromosome number within Odonata varies from 2n = 8 to 2n = 30, and the presumed ancestral karyotype, which is found in many lineages of Odonata, is 2n = 25. The sex-determination system of Odonates is XO but an XY system has evolved in 20 separate instances, likely from X-autosome fusions. Odonata chromosomes are usually monocentric, but holokinetic chromosomes appear to have evolved several times throughout Odonata.

However, despite previous research into Odonata chromosome evolution, many questions remain. While karyotype number can be used to draw evolutionary conclusions, the haploid number alone can fail to capture inferred changes when comparing chromosome-level genome assemblies in insects, failing to recover inversion, translocation, and sometimes even fission and fusion events. Thus, the amount of conservation between chromosomes within Odonata is not well understood. It is also unclear how repetitive elements and GC content influence the density of coding sequences or the degree of homology in a given area of Odonata chromosomes. Another enigma is the presence of micro-chromosomes ("m" chromosomes) throughout Odonata and dragonfly species. Ischnura elegans (2n = 25), while Pantala flavescens (2n = 23) and Tanypteryx hagenii (2n = 17) have a reduced karyotype, and Ischnura elegans (2n = 27) has expanded upon the ancestral karyotype. Here we compare and contrast the chromosome architecture of these four species of odonates and discuss this in the context of their evolutionary histories.
chromosome, and all four species use an XO sex determination system\textsuperscript{16}. The species are extremely variable in their natural histories. \textit{Platycnemis pennipes} and \textit{Ischnura elegans} are highly territorial odonates with relatively limited vagility, while \textit{Pantala flavescens} is considered the global wanderer, given their globally panmictic population, with individuals potentially crossing oceans\textsuperscript{22}. Further, these 3 species are all relatively young lineages from species rich families\textsuperscript{23,24}. Conversely, \textit{Tanypteryx hageni} is a member of the smallest family within Odonata\textsuperscript{25} estimated to have split from its sister species some 70 million years ago. Throughout its history, its morphology has remained largely unchanged designating it as a “living fossil”\textsuperscript{25}. Because of this, \textit{T. hageni} is of particular interest as a point of comparison to other long lived lineages, such as the coelocanth\textsuperscript{26}, lungfish\textsuperscript{27}, ginkgo tree\textsuperscript{28}, and mangrove horseshoe crab\textsuperscript{29}. Despite a small sample size, given the breadth of evolutionary history and life history traits represented, these Odonata genome assemblies are well-suited for answering outstanding questions regarding the chromosome evolution of Odonata.

**Results**

**Genome annotation**

As mentioned above, there are four long read chromosome assemblies available, representing the breadth of the evolutionary history of Odonata\textsuperscript{19} and we examined them all in a comparative context here. All are highly complete and contiguous\textsuperscript{19}. All but the genome of \textit{Platycnemis pennipes} had been previously annotated, so we annotated this genome using miniprot\textsuperscript{30} v.0.4, similar to the process used to annotate the genome of \textit{Tanypteryx hageni}\textsuperscript{5}. We recovered 22,810 protein coding genes from \textit{Platycnemis pennipes}. BUSCO\textsuperscript{31} analysis showed a high level of completion (C:91.7\% [S:89.2\%,D:2.5\%,F:2.5\%,M:5.8\%,n:1367]). This is comparable to other Odonata genome annotations, which have between 15,000 and 27,000 genes, with over 90\% of single copy orthologs present\textsuperscript{19}. Repeat content was widely variable in publicly available chromosome level Odonata genome assemblies, from approximately 18\% of the genome in the damselfly \textit{Pantala flavescens}, to over 50\% in \textit{Tanypteryx hageni} (Fig. 1). (See supplementary Fig. 1 for repeat analysis of all publicly available Odonata genome assemblies).

**Synteny analysis**

To identify homologous regions between the four species chromosomes, we compared the relative gene position and order. In a comparative genomic context, this is termed synteny; the conserved order of loci on the chromosomes of related species, which originated from a common ancestor\textsuperscript{32}. Despite being separated by 250 million years of evolution, the chromosomes of Odonata show clear homology (Fig. 2). Synteny between chromosomes demonstrates that \textit{Ischnura elegans}, \textit{Tanypteryx hageni} and \textit{Pantala flavescens} all have clear homologs to the ancestral karyotype (Fig. 2, Fig. 3), despite being separated by ~100 million years of evolution from \textit{Ischnura elegans} and over 250 million years from the two dragonflies\textsuperscript{24}. The karyotype of \textit{Ischnura elegans} expanded through one fission event, the chromosome number in \textit{Tanypteryx hageni} decreased through three fusion events, and the \textit{Pantala flavescens} chromosome number decreased through the loss of the m Chromosome (Fig. 2). The chromosomes resulting from fusion still show homology with the two ancestral chromosomes (Fig. 2).

There are only three instances of chromosomal regions sharing synteny with multiple chromosomes from the ancestral haplotype. Chromosome three in \textit{Tanypteryx hageni} (resulting from a fusion of chromosomes 10 and 11 in the ancestral haplotype) shares two syntenic regions with chromosome 11 in the arm of the chromosome that is syntenic with chromosome 10 (Fig. 3, 4). Chromosomes 9 and 12 in \textit{Ischnura elegans}, which resulted from the fission of chromosome 2 in the ancestral karyotype, each largely have synteny with only one half of chromosome 2 in \textit{Platycnemis pennipes} (Fig. 3, 4). However, chromosome 12 of \textit{Ischnura elegans} shares one syntenic region with the opposite chromosome arm of \textit{Platycnemis pennipes} (Fig. 3, 4). There are no other instances of a chromosome, or chromosomal arm in the case of chromosomes resulting from fusion, sharing synteny with more than one chromosome in \textit{Platycnemis pennipes} (Fig. 3). The overall structure of chromosomes has largely been maintained in these four species, with minimal influence of inversions, translocations, or even exogenous effects (such as transposable elements that can move between different regions of the genome and can reshuffle genes\textsuperscript{34}) on the relative gene order on each chromosome.

**Repetitive landscape**

Not only does the genome of \textit{Tanypteryx hageni} have a higher proportion of repeats in its genome than any other sequenced odonate genome (supplementary fig. 51), but the repeat landscape is also dramatically divergent from other odonates. While \textit{Ischnura elegans}, \textit{Platycnemis pennipes} and \textit{Pantala flavescens} have relatively constant levels of various repetitive elements throughout their repeat landscapes, the repetitive features in \textit{Tanypteryx hageni} appear in bursts (Fig. 5). This is also reflected in the genome, where transposons and retroelements especially appear in dense clusters throughout the chromosomes, and seem to be associated with a low coding sequence (CDS) content (Fig. 2). Repetitive elements have played an especially dramatic role in shaping the X chromosome of \textit{Tanypteryx hageni}, where unclassified (repetitive elements that did not have a match in the RepeatMasker\textsuperscript{35} library), transposable (elements that can physically move to other areas of the genome) and retroelements (elements that can “copy and paste” themselves to other areas of the genome through an RNA intermediate) are densely concentrated across long regions of the chromosome with a low CDS content and little synteny (Fig. 2).

**m chromosomes**

While the large autosomal chromosomes showed a high level of conservation, the micro chromosomes (defined here as chromosomes that were much smaller than all other chromosomes in the genome) did not. The m chromosomes of \textit{Platycnemis pennipes} and \textit{Tanypteryx hageni} shared one instance of synteny (Fig. 2), but when the genes on each m chromosome were BLASTed\textsuperscript{36} against all other protein coding sequences in the four genomes we found that they all shared a large number of orthologous genes (Fig. 6), indicating that they all likely descended from the same ancestral m chromosome. All three m chromosomes contain a number of intrachromosomal homologous genes, suggesting a high amount of duplication of the genes on these chromosomes, which could explain the lack of synteny. While the m chromosome of \textit{Pantala flavescens} is not present, the telomere of chromosome three of \textit{Pantala
flavescens does share one collinear block with the m chromosome of Tanypteryx hageni (Fig. 2), and this chromosome contains a number of orthologs to the m chromosomes of Ischnura elegans, Platycnemis pennipes, and Tanypteryx hageni (Fig. 6).

**Gene duplication**

Duplications do not appear to have occurred on a genome, or even chromosome, level but they have occurred on a smaller, locus scale. Ks plots (Fig. 7) comparing duplicate gene pairs show that gene duplicates are present in each genome, and are relatively recent in Pantala flavescens (mean Ks = 2.98, median Ks = .02), Ischnura elegans (mean Ks = 7.57, median Ks < .01) and Platycnemis pennipes (mean Ks = 9.68, median Ks = .01), but are likely much older in Tanypteryx hageni (mean Ks = 26.05, median Ks = 5.76).

**Modeling chromosome evolution**

To test for statistical signals for our observations, we partitioned each chromosome into linear 10 million base pair partitions (dropping the remainder at the end of each chromosome) and considered if the number of syntenic blocks in each partition was related to ancestral chromosome, chromosome type (micro, sex, or autosome), chromosome history (whether the partition was on a chromosome that resulted from fusion or fission), CDS content, GC content, and the proportion of unclassified repeats, retroelements, transposons and other repetitive elements. Given that the number of syntenic blocks is a count-based outcome, we used generalized linear models to model the number of syntenic blocks using the Poisson distribution. Since partitions from a given chromosome may be similar to each other in regards to the number of syntenic blocks, we also compared models with and without a random effect of chromosome using AIC values; analysis indicated random effects were not needed(supplementary table 1, supplementary Fig. 2). For each species-level model we used backwards selection procedures based on AIC values to determine final relationships to retain in the model37. Assumptions regarding dispersion and residuals were also checked for all final models, and we also evaluated the final fit using McFadden's pseudo-R² value38.

Platycnemis pennipes (McFadden's pseudo-R² = .14) and Ischnura pennipes (McFadden's pseudo-R² = .19) demonstrated lower fit than Pantala flavescens (McFadden's pseudo-R² = .29) and Tanypteryx hageni (McFadden's pseudo-R² = .34) (table 1). For McFadden's pseudo-R², values .2-.4 are considered good fits, matching traditional values of .7 − .939. Although some factors were found to impact the number of syntenic blocks in all species, none of the models retained the same subset of factors (table 1). The relative proportion of coding sequence (CDS content) was the only factor retained in all models and was found to be positively correlated with the number of collinear blocks per partition; this means that more coding sequence was correlated with more syntenic blocks. While it is clear that gene duplication has occurred (Fig. 6), it does not appear to have had a major impact upon synteny.

The corresponding ancestral chromosome was retained in all species but Platycnemis pennipes. Chromosomes 6,7,9,10,11,X and m had a negative regression coefficient in T. hageni, I. elegans, and P. flavescens (supplementary table 2). The directionality of the coefficients of the other chromosomes varied by species (supplementary table 2). In Platycnemis pennipes, where the chromosome classification was retained and corresponding ancestral chromosome was not, both sex and micro chromosomes had a negative regression coefficient (supplementary table 2).

The effect of various repetitive elements varied across taxa. Higher proportions of retroelements significantly negatively influenced the number of collinear blocks in Tanypteryx hageni and Ischnura elegans (supplementary table 2). Notably, both species displayed a relatively low proportion of retroelements of the class SINE (short interspersed nuclear elements) (supplementary fig S2). Unclassified repeats were negatively correlated with synteny in Platycnemis pennipes (supplementary table S2), but positively influenced synteny in Ischnura elegans. Transposons had a negative influence upon synteny in Ischnura elegans, while all other repeats were significantly negatively correlated with synteny in Ischnura elegans, and Platycnemis pennipes (supplementary table S2).

Relative position was only retained in Ischnura elegans, where partitions towards the terminal end of the chromosome had fewer syntenic blocks on average. Tanypteryx hageni was the only species for which GC content was retained, and it was negatively related to synteny.

Because retroelements play an outsized role in the genomes of other long lived lineages37,38, and retroelements were retained in the final models generated for Tanypteryx hageni and Ischnura elegans, we replaced the overall proportion of retroelements with the proportion of three categories of retroelements (long interspersed nuclear elements/LINEs, which included elements identified as PENEOPE/LINE by repeatmasker, short interspersed nuclear elements/SINEs, and long terminal repeats/LTRs) and used backwards selection procedures based on AIC values to determine which type of retroelements were leading to observed relationship. We again evaluated the final model fit using McFadden's pseudo-R² value.

LINEs were retained in both Tanypteryx hageni and Ischnura elegans (table 2), and had a negative correlation coefficient in both species (supplementary table 3). SINEs were also retained in I. elegans (table 2) with a positive regression coefficient (supplementary table 3). Long-terminal repeats were not retained for either species-level model.

**Discussion**

The level of conserved autosomal synteny shown in these four species of Odonata is comparable to that seen in Aves8–11 and Lepidoptera15, yet Anisoptera and Zygoptera diverged from each other 185 my and 110 my before the radiation of Aves and Lepidoptera, respectively. In the 250 million years since the split between Anisoptera and Zygoptera, there is scant evidence for any translocation events, although fusion and fission appear to be somewhat common16.

In these Odonata species, a higher CDS content is a strong predictor of synteny (table 1). It is plausible that local gene duplication events in CDS rich regions could disrupt synteny, and although there are measurable amounts of gene duplications in each genome (Fig. 6), this does not seem to be influencing synteny in any measurable way. Repetitive elements (table 1), are largely negatively correlated with synteny, presumably because they are displacing CDS.
Chromosomes in *Tanypteryx hageni* that originated from a fusion event did not show less synteny than other autosomes (table 1), but chromosome 9 of *Ischnura elegans*, which is the product of a fission event, did have significantly less synteny than other autosomes. It is possible that synteny is lost in chromosomes resulting from fission in Odonata, but a much larger sample size would be needed to make a definitive conclusion. Statistical model showed that chromosomes sharing synteny with chromosomes 6,7,9,10,11,X and m in *Platycnemis pennipes* had a negative regression coefficient in *T. hageni*, *I. elegans*, and *P. flavescens*. These regions tend to have noticeable gaps in synteny as mapped in the circos plot (Fig. 1).

Contrary to their static autosomal counterparts, the sex and micro chromosomes of Odonata are fluid. Partitions from the sex chromosomes (table 1) had fewer syntenic blocks than autosomes on average. Given the relative stability of autosomes, turnover on the X chromosome could be an important driver of speciation in Odonata. It is understood that additions and deletions can accumulate more rapidly on sex chromosomes, and that sex chromosomes can contain numerous selfish elements that influence genome stability and can quickly lead to reproductive barriers and speciation. There is a possibility that the species within the family Petaluridae, including *Tanypteryx hageni* have a strict, multi-year larval period, which could potentially lead to differential population structure between cohorts. The especially high accumulation of transposable elements on the X chromosome of *Tanypteryx hageni* is a plausible reproductive barrier between different cohorts. Further research is needed to investigate this phenomenon.

Because the micro chromosomes of *Platycnemis pennipes*, *Tanypteryx hageni* and *Ischnura elegans* share a large number of potentially homologous genes (Fig. 5), it is likely they descended from an ancestral micro-chromosome. Partitions from the m chromosomes of *Tanypteryx hageni* and *Platycnemis pennipes* contained less synteny than autosomes (supplementary table 2). Clearly there are fewer selective pressures to maintain the chromosomal structure of the m chromosomes in these Odonata. A telomere of the third chromosome of *Pantala flavescens* shares a large number of potentially homologous genes with all three micro chromosomes, and one syntenic block with the m chromosome of *Tanypteryx hageni*, suggesting a possible fusion event that resulted in the loss of the m chromosome. We acknowledge a larger sample size of Odonata genomes is needed to determine the fate of missing m chromosomes, and confirm if all m chromosomes across the order are homologous.

The overall trends in the micro chromosomes of Odonata are much different than other lineages within the tree of life. In birds, microchromosomes contain a large number of genes, are highly conserved, and contain few transposable elements. Many m chromosomes, however, have been lost in reptiles and mammals. In the Australian Giant Lungfish the microchromosomes also have a high CDS content, and a particularly low concentration of LINEs compared to the rest of the genomes. Further research is needed to determine if the fluid, repeat heavy micro chromosomes of *Platycnemis pennipes*, *Tanypteryx hageni*, and *Ischnura elegans* are unique to Odonata, or if they can be found more commonly throughout insecta, or even arthropoda.

Our generalized linear models had a better fit in the two anisoptera (McFadden's pseudo-R² = .29,.34) than the zygoptera (McFadden's pseudo-R² = .14,.19). This could be due to differential processes shaping the genomic evolution in each suborder. However, given our sample size it would be inappropriate to make such a conclusion. As more chromosome level assemblies from Odonata differences between the suborders can be more thoroughly explored.

When compared to the other publicly available genome assemblies from Odonata, *Tanypteryx hageni* has an unusually high repeat content (Fig. 1, supplementary fig. S1), which has resulted in a larger genome size than any other Odonata for which there is a genome assembly. It also has a repeat profile showing punctuated instances of repetitive element activity (Fig. 5), which could very well result in large areas of insertions on chromosomes, as opposed to the relatively smooth equilibrium seen in the other three species. Retroelements are the only repetitive elements found to disrupt synteny in the species (table 1) so it is plausible that retroelements generally, and LINEs in particular (table 2) are playing an outsized role in shaping the genome of *Tanypteryx hageni*.

The genome assemblies of other so-called “living fossils” also contain unique repeat signatures. The genome of the gingko biloba tree is primarily composed of LINEs, which could have similarly influenced repeat content. Although it is phylogenetically distant from *Tanypteryx hageni*, the Giant Australian Lungfish (*Neoceratodus forsteri*) has a similar evolutionary history. Like *Tanypteryx hageni*, *Neoceratodus forsteri* has a small population size, presumably for quite some time. It has not undergone a whole genome duplication event since the diversification of vertebrates, and it diverged from sister species ~ 190 MYA. Both *Tanypteryx hageni* (supplementary fig. S2) and *Neoceratodua forsteri* have a high proportion of LINEs in their genomes, and LINEs were retained in the model retroelement model generated for *Tanypteryx hageni*. In *N. forsteri*, it has been demonstrated that the activity of these elements has resulted in an ever growing size. It is quite possible that in both lineages, prolonged periods of a low effective population size led to the accumulation of LINEs. *Neoceratodua forsteri* does have a much larger proportion of LINEs than *Tanypteryx hageni* (~ 25% in *N. forsteri* and 12.37% in *T. hageni*). This could be due to phylogenetic differences or the older species age of *Neoceratodua forsteri*.

Our models are an important step towards understanding the genomic evolution of Odonata. However, as more genomes become publicly available further investigation should be undertaken to how various genomic elements interact with each other, and perhaps how different classification and annotation methods could impact findings.

**Methods**

**Dataset**

To determine the amount of conserved synteny across Odonata we used the chromosome level genome assemblies of the Zygoptera *Platycnemis pennipes* (downloaded from NCBI) and *Ischnura elegans* and the Anisoptera *Pantala flavescens* and *Tanypteryx hageni*. All genomes have been shown to be of
Relatively high quality\(^\text{19}\), and all but \textit{P. pennipes} have been annotated. We used the remaining Odonata genome assemblies from NCBI (\textit{Rhinocypha anisoptera}, \textit{Ladona fulva}\(^\text{45}\), \textit{Calopteryx splendens}\(^\text{46}\), and \textit{Hataerina americana}) to compare repeat content across the order.

**Genome annotation**

Repetitive elements for all publicly available Odonata genome assemblies were modeled using RepeatModeler v2.0.1\(^\text{47}\) and classified with RepeatMasker2 v4.1.2\(^\text{47}\). Prior to running RepeatModeler, a BLAST\(^\text{36}\) database was created for each genome using BLAST\(^\text{37}\) v2.12.0. RepeatModeler was then run on the BLAST database to identify repetitive sequences. Following this, RepeatMasker was run using the option \texttt{--nolow}, which does not mask simple repeats. RepeatMasker then classified the modeled repetitive elements into families and provided a classification summary for each genome.

To annotate the genome of \textit{Platycnemis pennipes} we clustered the protein sets of \textit{Ladona fulva}\(^\text{45}\), \textit{Ischnura elegans}\(^\text{31}\), \textit{Pantala flavescens}\(^\text{20}\), and \textit{Tanypteryx hageni}\(^\text{18}\) to 40 percent similarity with CD-HIT v4.8.1\(^\text{48}\). To ensure we were using a high quality reference set, we ran BUSCO\(^\text{31}\) on the clustered protein set, using the insecta database (odb10) as a reference, and recovered 94.9% of the single and complete BUSCO genes, with an additional 2.7% duplicated and 0.7% fragmented.

We then mapped the clustered reference set onto the masked assembly of \textit{Platycnemis pennipes} using miniprot v0.4\(^\text{30}\). We extracted the protein set from the resulting gff file using gffread\(^\text{49}\), and assessed the annotation completion by running BUSCO on the annotated protein set, with the BUSCO insecta database odb10 as a reference\(^\text{31}\).

**Syntenic analysis**

To visualize synteny between \textit{Tanypteryx hageni}, \textit{Pantala flavescens}, \textit{Ischnura elegans} and \textit{Platycnemis pennipes} we first identified potential homologous genes between species using BLASTp\(^\text{36}\) to compare amino acid sequences from each genome annotation to a custom database containing all amino acid sequences from the three other genomes, limiting output to the top five hits between two species per gene, and restricting output to an e-value below 1e-5. We then combined the BLAST output and gff annotation files for each species and used MCScanX\(^\text{50}\) to identify areas with conserved synteny.

We visualized conserved syntenic blocks between the chromosomes of each species, highlighting synteny with the ancestral karyotype conserved in \textit{Platycnemis pennipes} in a bar plot generated with VGSC\(^\text{51}\). We also generated a line plot, with collinear portions of the chromosome plotted alongside the corresponding proportion of coding sequence (CDS) (\textit{min} = 0.0, \textit{max} = 0.13), GC content (\textit{min} = .30, \textit{max} = 0.43), retroelements (\textit{min} = 0.0, \textit{max} = .65), transposons(\textit{min} = 0.0, \textit{max} = .65), unclassified repetitive elements(\textit{min} = 0.0, \textit{max} = 0.65), and all other repetitive elements(\textit{min} = 0.0, \textit{max} = .65) using Circos v.0.69–18\(^\text{52}\). All repetitive element proportions were calculated from repeatmasker2 output\(^\text{47}\).

To highlight chromosomal fission and fusion, we generated a second Circos plot\(^\text{53}\) displaying only the involved chromosomes, as identified by the first plot, and the syntenic blocks between them.

**M chromosome links**

To further investigate the origins of the micro chromosomes of \textit{Tanypteryx hageni}, \textit{Ischnura elegans}, and \textit{Platycnemis pennipes} we used BLASTp\(^\text{36}\) to identify potentially homologous genes between genes located on the m chromosomes, and the entirety of all four genomes. We generated a custom database using the translated amino acid sequences from all four genome annotations, and limited the number of blast hits per gene to five, and restricted output to an e-value below 1e-5. We visualized the links, and their locations on each chromosome using Circos v.0.69–18\(^\text{53}\).

**Repeat Landscape**

We generated a repeat landscape for each of the four chromosome-level assemblies by parsing the repeatmasker output using the script parseRM_GetLandscape.pl from the perl module Parsing-RepeatMasker-Outputs\(^\text{52–55}\). We used ggplot2\(^\text{56}\) in r to visualize the whole repeat landscape for each species. Elements that were classified as PENELPOES/LINEs by repeatmasker were classified as LINEs (long interspersed nuclear elements) in the repeat landscapes.

**Ks plots**

To visualize the relative age of duplication events in \textit{Platycnemis pennipes}, \textit{Ischnura elegans}, \textit{Tanypteryx hageni}, and \textit{Pantala flavescens}, we uploaded each genome and its corresponding gff file to CoGe\(^\text{57}\). We then used the synmap2 function\(^\text{58}\) to identify genomic paralogous pairs, using the option to "calculate syntenic CDS pairs and color dots" with synonymous substitution rates to generate a Ks plot for each species.

**Statistical analysis**

To test for statistical signals for our observations, we partitioned each chromosome into linear 10 million base pair partitions (dropping the remainder at the end of each chromosome) and considered if the number of syntenic blocks in each partition was related to ancestral chromosome, chromosome type (micro, sex, or autosome), chromosome history (whether the partition was on a chromosome that resulted from fusion or fission), CDS content, GC content, and the proportion of unclassified repeats, retroelements, transposons and other repetitive elements. We first considered if potential relationships between these factors and the number of syntenic blocks differed among species by constructing models that included main effects of each of these factors and main effects of species (suggesting impacts of factors differed among species but in an additive fashion), no effect of species (suggesting impact of these factors did not depend on species), and both main effects of species and interactions with these factors (suggesting impact of these factors differed among species).
Given that the number of syntenic blocks is a count-based outcome, we used generalized linear models to model the number of syntenic blocks using the Poisson distribution. Since partitions from a given chromosome may be similar to each other in regards to the number of syntenic blocks, we also compared models with and without a random effect of chromosome using AIC values\textsuperscript{24} for both model types, the lowest AIC score was observed when interactions between species and other factors were included, indicating relationships differed among species. The lowest overall AIC score was associated with models without random effects, thus we developed these models for each species for further investigation. For each species, we used backwards selection procedures based on AIC values to determine final relationships to retain in the model\textsuperscript{37}. Assumptions regarding dispersion and residuals were also checked for all final models, and we also evaluated the final fit using McFadden's pseudo-$R^2$ value\textsuperscript{38}.

**Declarations**

**Competing interests**

The authors declare no competing interests.

**References**


41. Baird, I. The wetland habitats, biogeography and population dynamics of Petalura gigantea (Odonata: Petaluridae) in the Blue Mountains of New South Wales. (The University of Western Sydney, 2012).


Table One: Factors retained in final model for each species

<table>
<thead>
<tr>
<th>Species</th>
<th>Mcfadden's pseudo-R²</th>
<th>Factors retained</th>
<th>Relative proportion of:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ancestral chromosome</td>
<td>chromosome classification</td>
</tr>
<tr>
<td>P. pennipes</td>
<td>.14</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>I. elegans</td>
<td>.19</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>P. flavescens</td>
<td>.29</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>T. hageni</td>
<td>.34</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

Table one: Output from generalized linear models using chromosome classification (sex chromosome, micro chromosome, or autosome), corresponding ancestral chromosome, chromosome state change (fusion, fission, or none), CDS content, GC content, relative position of the partition on the chromosome, and the proportion of unclassified repeats, retroelements, transposons and other repetitive elements influenced the number of syntenic blocks in chromosome partitions for T. hageni, I. elegans, and P. pennipes.

Note: State change was not retained in any species, and was not an option for P. pennipes or P. flavescens.

x=retained in model.

Table two: Factors retained in retroelement model for each species

<table>
<thead>
<tr>
<th>Species</th>
<th>Updated McFadden's pseudo-R²</th>
<th>Original McFadden's pseudo-R²</th>
<th>Factors retained</th>
<th>Relative proportion of:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>ancestral chromosome</td>
<td>chromosome classification</td>
</tr>
<tr>
<td>I. elegans</td>
<td>.21</td>
<td>.19</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>T. hageni</td>
<td>.35</td>
<td>.34</td>
<td>x</td>
<td></td>
</tr>
</tbody>
</table>

Table one: Output from generalized linear models using retained factors from original models, with retroelements broken down into the relative proportion of long interspersed nuclear elements, short interspersed nuclear elements, and long terminal repetitive elements, for species where retroelements were retained in the original model.

Note: Long terminal repeats were not retained in either model.

x=retained in model.

Figures
Figure 1

Repetitive content of chromosome-level genome assemblies from Odonata

Data on repetitive content of four chromosome-level, long-read Odonata genomes. The relative proportion of retroelements, DNA transposons, rolling-circles and unclassified repetitive elements are compared.

Figure 2

Synteny in Odonata genomes

A) Configuration of chromosomes of P. pennipes, which has retained the ancestral karyotype. Chromosomes of B) I. elegans, C) P. flavescens and D) T. hageni, with regions of synteny to P. pennipes chromosomes color coded to match the chromosomes from the ancestral karyotype.
Figure 3

Comparative plot of Odonata genomes

Circos\textsuperscript{23} plot of all four chromosome level Odonata genome assemblies. Lines connect inter-species syntenic regions. Synteny with P. pennipes is color coded according to the color of the P. pennipes chromosome. All other syntenic links are in grey. This is followed by the relative proportion of CDS in blue (min=0, max=.15), and the relative GC content in red (min=.30, max=0.43). The next track displays the chromosomes from each genome (Plat=P. pennipes, Tan=T. hageni, Pan=P. Flavescens, and Isch=I. elegans). Retroelements (red), transposons (blue), unclassified repeats (purple), and all other repeats (black) are displayed on the outer track (min=0, max=.65)
Mismatched syntenic blocks

*Figure 4*

A [Circos](https://circos.ca) plot showing chromosomes with mismatched syntenic links. Chromosomes 9 and 12 are displayed from *Ischnura elegans (Isch)*, chromosome 3 is displayed from *T. hageni (Tan)*, and chromosomes 9, 11, and 12 are displayed from *P. pennipes (Plat)*.
Repeat Landscapes of Odonata Genomes

Repeat landscapes of A) *P. pennipes*, B) *I. elegans*, C) *P. flavescens*, and D) *T. hageni*. For each species the Kimura substitution level of the repetitive elements is shown on the x axis as a proxy for age. The proportion of the genome for elements with a given Kimura substitution level is shown on the y axis. Varying repetitive elements are colored according to the legend.
Figure 6

BLAST hits of micro chromosomes

*Links between genes on the m chromosomes of T. hageni, I. elegans, and P. pennipes and significant BLAST hits throughout all four genomes.*
Figure 7

Ks plots of Odonata genomes.

Displays the number of synonymous substitutions, as a substitute for time, between a paralogous gene pair on the x-axis, and frequency on the y-axis for A) P. pennipes B) I. elegans, C) P. flavescens, and D) T. hageni. Note the differing tscales for each species.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementarymaterials.docx