scientific reports

OPEN

Indel driven rapid evolution of core nuclear pore protein gene promoters

David W. J. McQuarrie[1,2,4], Adam M. Read[1,4], Frannie H. S. Stephens[1], Alberto Civetta[3✉] & Matthias Soller[1,2✉]

Nuclear pore proteins (Nups) prominently are among the few genes linked to speciation from hybrid incompatibility in *Drosophila*. These studies have focused on coding sequence evolution of Nup96 and Nup160 and shown evidence of positive selection driving nucleoporin evolution. Intriguingly, channel Nup54 functionality is required for neuronal wiring underlying the female post-mating response induced by male-derived sex-peptide. A region of rapid evolution in the core promoter of *Nup54* suggests a critical role for general transcriptional regulatory elements at the onset of speciation, but whether this is a general feature of Nup genes has not been determined. Consistent with findings for *Nup54*, additional channel Nup58 and Nup62 promoters also rapidly accumulate insertions/deletions (indels). Comprehensive examination of Nup upstream regions reveals that core Nup complex gene promoters accumulate indels rapidly. Since changes in promoters can drive changes in expression, these results indicate an evolutionary mechanism driven by indel accumulation in core Nup promoters. Compensation of such gene expression changes could lead to altered neuronal wiring, rapid fixation of traits caused by promoter changes and subsequently the rise of new species. Hence, the nuclear pore complex may act as a nexus for species-specific changes via nucleo-cytoplasmic transport regulated gene expression.

The nuclear pore protein complex (Nup complex) provides a physical barrier between the nucleus and cytoplasm requiring active transport for cargos above about 40 kDa[1]. The Nup complex consists of 30 proteins conserved in *Drosophila* (Fig. 1), which are grouped into subcomplexes termed outer ring (OR), inner ring (IR), cytoplasmic filaments (CF), nuclear basket (NB) and pore membrane proteins (POMS). The transport channel is made up of three phenylalanine-glycine-rich (FG) repeat domains of Nup54, Nup58 and Nup62. The Nup complex can contribute to differential expression of other genes as e.g. actively transcribed genes can be in the proximity of the Nup complex[2–5].

Male derived substances transferred during mating change female physiology and behaviour to guarantee reproductive success in many insects[6–9]. Rapid divergence in male versus female interactive molecules can hamper reproductive success, potentially leading to the establishment of new species for example by compromising fertility or viability among hybrids[10–18]. Such hybrid incompatibility among closely related *Drosophila* species has been used to identify responsible genes. Among the few genes identified were nuclear pore proteins Nup96 and Nup160[19–22]. Interestingly, Nups are overrepresented among speciation genes[19]. In particular, Nup96 and Nup160 have been shown to be under adaptive evolution and at an individual gene level can cause hybrid lethality, contrasting with reports of multiple linked factors being necessary for hybrid lethality[19–21].

Sex peptide is the master regulator of the female post-mating response in *Drosophila melanogaster*[8,23]. A screen for sex-peptide insensitive mutants identified Nup54 functionality to be important for neuronal wiring of circuits involved in regulating post-mating behaviours[15]. In particular, a deletion in the promoter of the *Nup54* gene has been associated with altered nucleo-cytoplasmic shuttling in eight *pickpocket* (*ppk)* expressing neurons in the central brain leading to wiring defects and a compromised female post-mating response directed by male-derived sex-peptide transferred during mating[15]. Moreover, this *Nup54* promoter deletion allele maps to a region for rapid evolution in the *Nup54* promoter suggesting sexual conflict driving female escape from male manipulation by sex-peptide under unfavourable conditions[15]. Furthermore, channel Nups have also been attributed a role in

[1]School of Biosciences, College of Life and Environmental Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK. [2]Birmingham Centre for Genome Biology, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK. [3]Department of Biology, University of Winnipeg, Winnipeg, MB R3B 2E9, Canada. [4]These authors contributed equally: David W. J. McQuarrie and Adam M. Read. ✉email: a.civetta@uwinnipeg.ca; m.soller@bham.ac.uk
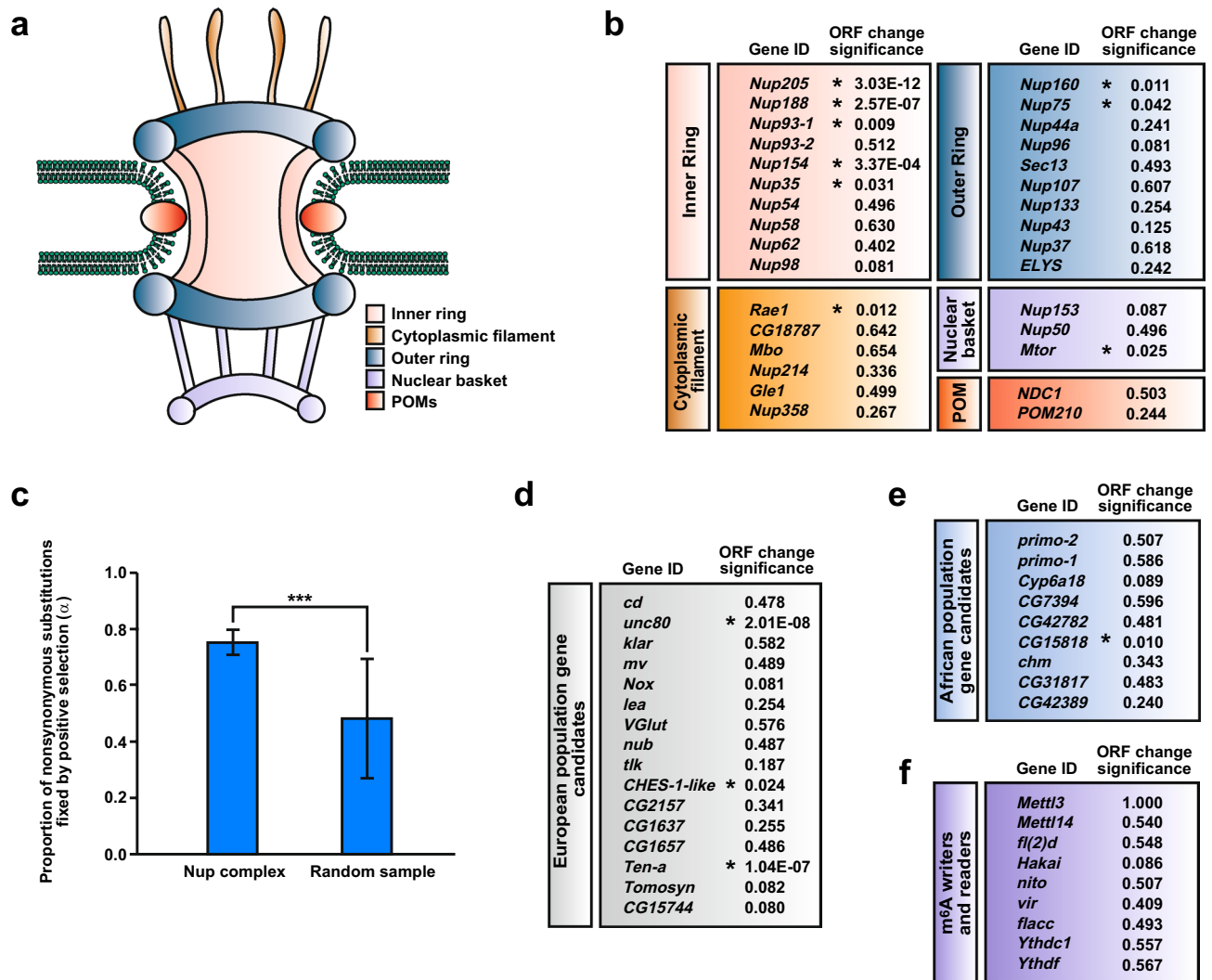
Figure 1. Positive selection drives the evolution of the nuclear pore complex proteins. (**a**) Graphic depiction of the nuclear pore complex with differentially coloured subcomplexes inner ring (IR, pink), outer ring (OR, blue), cytoplasmic filament (CF, orange), nuclear basket (purple) and pore membrane proteins (POMS, red) embedded in the nuclear envelope lipid-bilayer (green). (**b**) Significance of synonymous to non-synonymous changes in Nup subcomplex open reading frames (ORFs) indicated by an asterisk and p values. (**c**) Average proportion of substitutions fixed by positive selection (α) shown with standard deviation. 30 genes were randomly sampled 100 times with replacements from the list of Nups and from the genome. Statistically significant differences from unpaired student t tests are indicated by asterisks (*** $p \leq 0.0001$). (**d–f**) Significance of synonymous to non-synonymous changes in ORFs indicated by an asterisk and p values. As positive control groups the "best candidate genes" under balancing selection for European (**d**) and African (**e**) populations were used. As a negative control group, the highly conserved m6A writer complex and m6A readers were analysed (**f**).

transposon silencing in the germline by facilitating processing of short piRNAs from long pre-curser RNAs of the *flamenco* locus, which functions as a 'master off switch' for transposons[24], indicating that the disruption to Nup complex regulation likely has pleiotropic effects.

Given prior findings about adaptive evolution of the Nup96 and Nup160 coding region[19–22], we systematically extended this analysis to all Nups. Moreover, since the *Nup54* gene promoter evolves rapidly we systematically analysed all Nup gene promoter regions. Using *Drosophila* phylogenomics we reveal that promoters of core Nup complex genes are accumulating insertions/deletions (indels) driving rapid evolution. These findings suggest that the nuclear pore complex is a nexus for species-specific changes via nucleo-cytoplasmic transport regulated gene expression.

Results

Positive selection drives the evolution of nuclear pore complex proteins. To assess whether the protein coding region (Open Reading Frame, ORF) of Nups were under selection, we performed McDonald-Kreitman tests (MKTs) to compare the number of polymorphisms in the *D. melanogaster* ancestral Congo population, and divergence between *D. melanogaster* and its closest relative *D. simulans*. This analysis showed

positive selection for only nine Nup ORFs (Fig. 1b). Moreover, nucleoporins have a higher proportion of non-synonymous substitutions fixed by positive selection (α) (Fig. 1c, Supplementary Dataset 1)[25], meaning stronger evidence of positive selection, than expected from a random genome sample of equal size.

To substantiate the significance of the Nup complex subcomplex MKT results, we carried out the same analysis for additional sets of genes that either evolve rapidly or are highly conserved. We used two positive control groups identified in Croze et al. (2017)[26] as the "best candidate genes" under balancing selection for European (Fig. 1d) and African (Fig. 1e) populations. As a negative control group we analysed the m6A mRNA methylation machinery (Fig. 1f, writer complex: Mettl3, Mettl14, fl(2)d, virilizer, flacc, nito and Hakai; readers: YTHDC1 and YTHDF), because high evolutionary conservation is required to maintain complex stoichiometry to guarantee functionality making this group optimal for use as a control group to monitor protein evolution (Fig. 1f)[27–31]. Interestingly, the two positive control groups had a low number of rapidly evolving members for both the European (three out of sixteen) and African (one out of nine) groups (Fig. 1d, e). As expected, all m6A group members followed the neutral hypothesis (Fig. 1f). Taken together, these results demonstrate that Nup complex members are fast evolving, particularly the inner ring subcomplex (Fig. 1b).

Promoter regions of channel Nups 58 and 62 have diverged in closely related species.
Since we previously identified the promoter of the channel Nup54 as a region of rapid sequence divergence[15], we examined the promoters of the other two channel Nups 58 and 62 among closely related species of *Drosophila* (*D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba* and *D. erecta*) (Fig. 2). This detailed analysis revealed that also the promoters of the other two channel Nups accumulated indels rapidly (Fig. 2). Moreover, these changes contain a number of indels (Fig. 2c, d) that can fundamentally impact on transcription factor binding commonly found in the proximal region before the TATA box[32–36].

Next, we extended this detailed analysis to Nup98-96 and Nup160, since these have been implicated as drivers of speciation[19]. As with Nup58 and Nup62, this analysis indicated divergence in promoter regions of the two genes (Supplementary Fig. 1a, b) and highlighted accumulation of indels in promoters of Nup98-96 and Nup160 (Supplementary Fig. 1c, d).

Inner and outer ring nuclear pore protein gene subcomplexes undergo rapid evolution through indel accumulation in promoters.
Next, we examined the promoter region of all remaining Nups in the closely related *Drosophila* species to see whether accumulation of mutations is a general feature of the promoters of this class of genes (Fig. 3). We analysed conservation in 27 insect species through PhyloP27way data[37,38]. To determine whether the Nup complex evolves rapidly compared to the genome, we used PhyloP scores as a measure of conservation upstream of the transcription start site (TSS) (Fig. 3a). We observed a significant decrease in Nup complex sequence conservation in the promoter region upstream of the predicted TATA box site (−30 to −380), where the average PhyloP conservation score was significantly lower for the Nup complex compared to the genome (Fig. 3a, b). Additionally, accumulation of changes in promoter regions were analysed as PhyloP sliding window diversity scores (d^P) and compared between the Nup complex and the genome (Fig. 3c). Here, the Nup complex significantly accumulated sequence changes compared to the genome (Fig. 3c).

To focus more specifically on *Drosophila*, we analysed promoter evolution between species in the *melanogaster* subgroup[39]. As a control group we used the m6A mRNA methylation machinery control group again because they also have similar TSS properties to Nup complex genes comparative to other gene groups like ribosomal genes[40,41], making this group optimal for use as a control group to monitor promoter evolution[27–30].

We analysed accumulation of changes in promoter regions using sliding window diversity scores (*d*) and compared scores between Nup complex subcomplexes and the control group. As has been observed for *Drosophila* and human promoters[36,40,42], the region immediately before the TATA box constitutes transcription factor binding sites (−30 to −380), and showed a slightly increased rate of sequence changes in m6A mRNA methylation writer complex and readers consistent with the general trend of greater sequence diversity upstream of gene TSSs (Fig. 3d–i). We then analysed the promoters of the different Nup sub-complexes IR, OR, CF and NB. We investigated all sequence differences (Fig. 3d, e) and performed a more detailed analysis split into indels (Fig. 3f, g) and base changes (Fig. 3h, i). This analysis revealed that the core nuclear complex consisting of the IR and OR genes showed a significantly enhanced rate of indel driven evolution compared to the control group or the CF and NB group of genes (Fig. 3f, g). Analysis of all sequence differences mirrored the indel analysis including a significant increase in the number of differences in IR and OR Nups (Fig. 3d, e). In contrast base changes were reduced for OR and CF Nups (Fig. 3h, i).

Discussion
Here, through a combined analysis of polymorphism and divergence between closely related *Drosophila* species we show that the evolution of Nup complex proteins is driven by positive selection. These results are in agreement with previous findings of adaptive evolution of Nups[19]. Our MKT results support previously reported evidence of positive selection in these same species for Nup160, with only marginally significant evidence for Nup96[20,21]. Slight differences in results are not unexpected when different source populations are used for the analysis[16].

Focussing on upstream and downstream Nup gene regions, we reveal a bias for inclusion of nucleotide changes in the core members of the Nup complex. Consistent with channel *Nup54*[15], further analysis of upstream regions of the other channel Nups 58 and 62 reveals that their promotor regions rapidly accumulate indels. Through examination of promoter regions of Nup subcomplex members, we identify the core IR and OR Nups as rapidly accumulating indels. Here, the core Nup complex members likely act as drivers of species-specific variation through indel driven changes in promoters, subsequently modifying regulation of nucleo-cytoplasmic transport regulated gene expression.
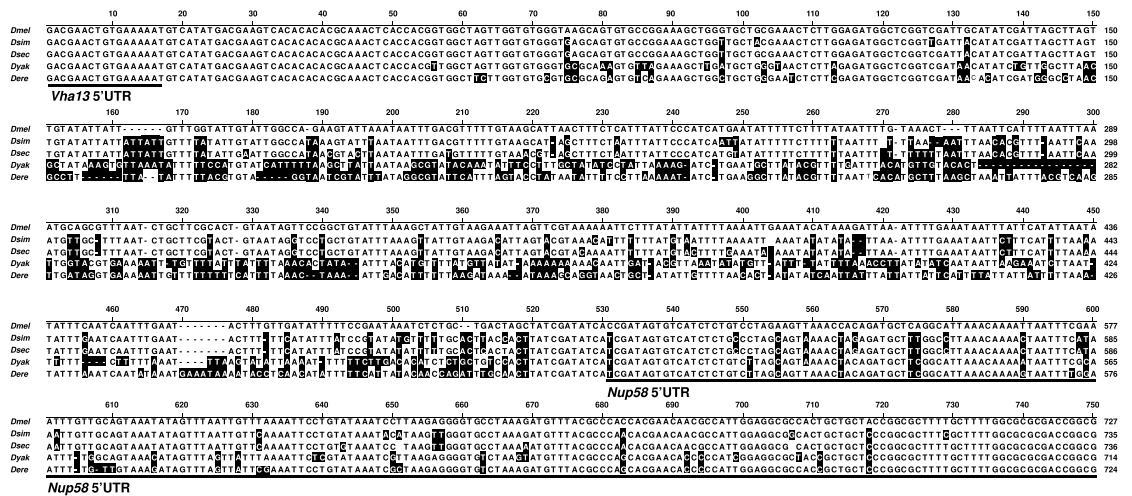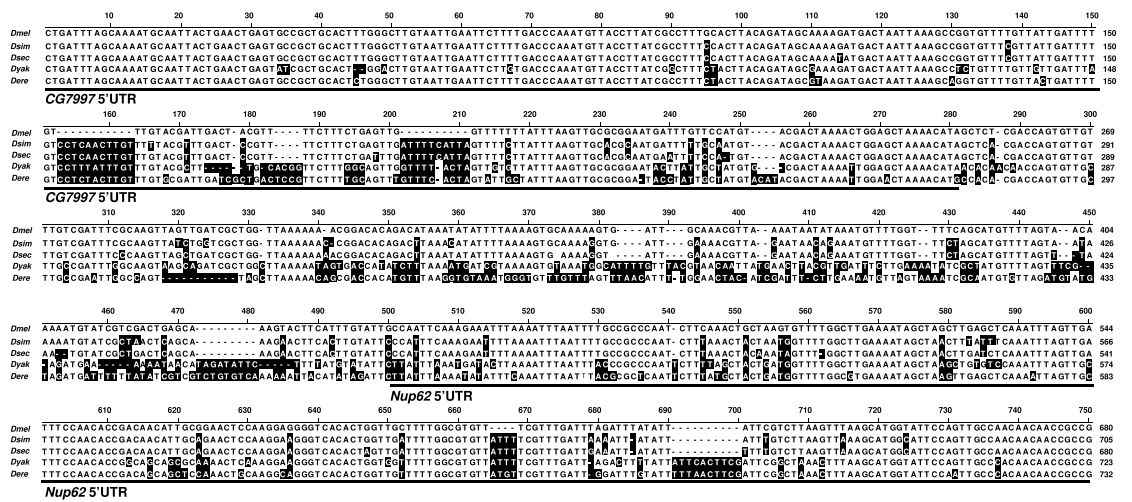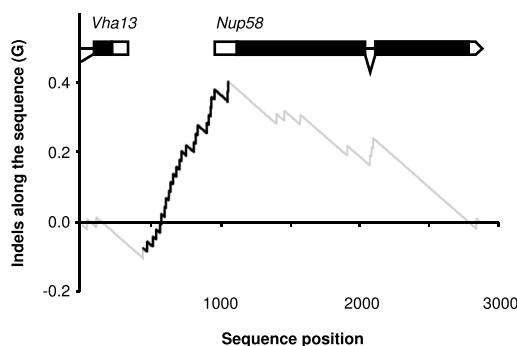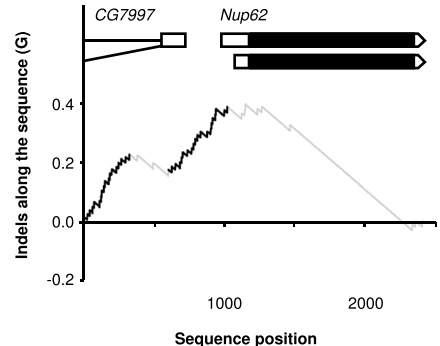
Figure 2. Promoter regions of channel Nups 58 and 62 have diverged in closely related species. **(a,b)** Sequence alignment of the *Nup58* and *Nup62* promoter regions from closely related species. Nucleic acids changes from *D. melanogaster* are indicated in black. Transcribed parts of the *Vha13* and *CG7997* 5′UTR and the *Nup58* and *Nup62* 5′UTR are indicated by a line. **(c,d)** Plot of cumulative differences along the sequence (G) between the relative occurrences of indels and their position from the alignment of the gene region around *Nup58* and *Nup62* between *D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba* and *D. erecta*. Positions in the alignment with significant stretches of substitutions are identified by black line(s).
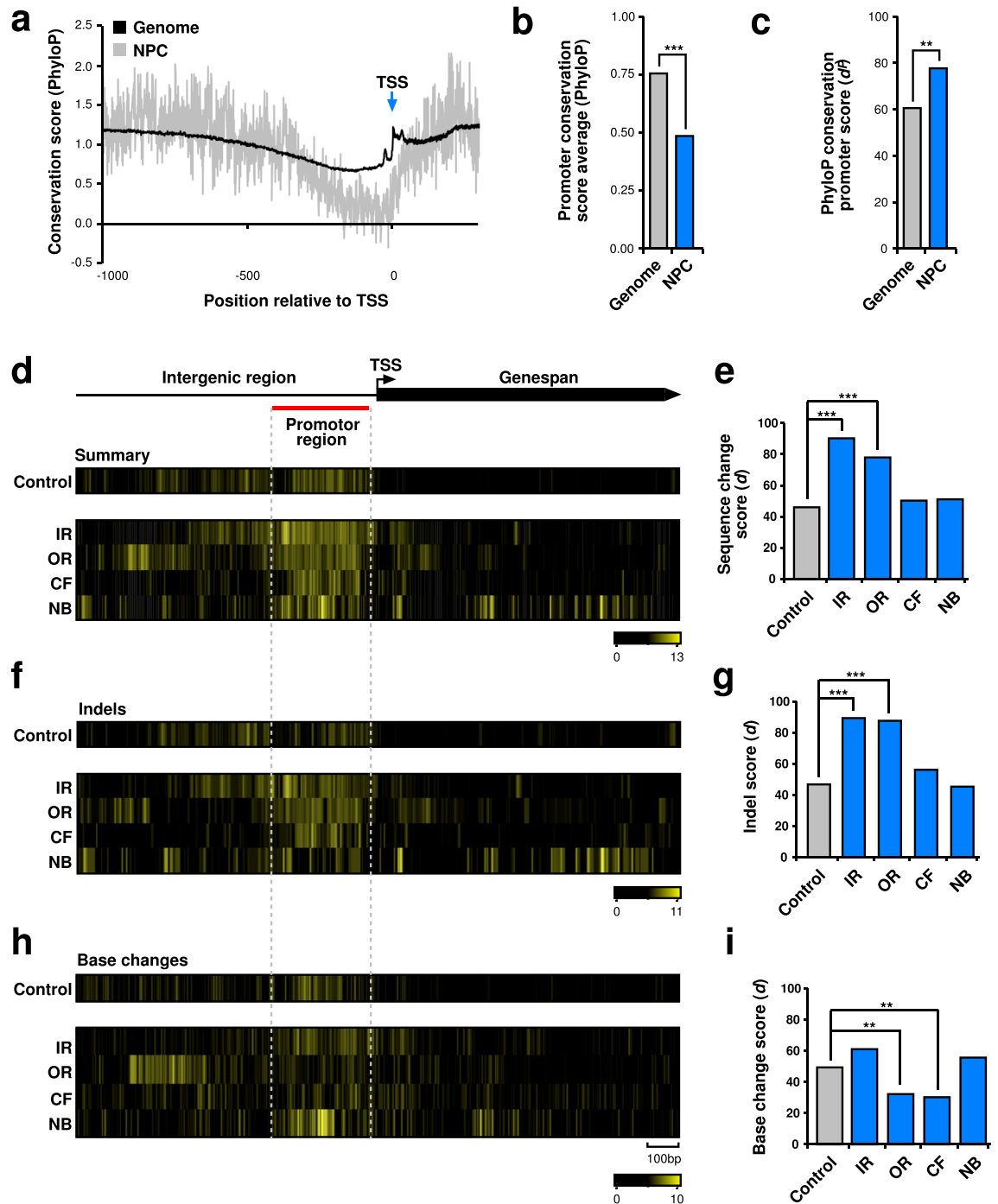
Figure 3. Core nuclear pore protein genes rapidly evolve through indel accumulation in promoters. (**a**) Evolutionary conservation of nucleotide positions around the TSS depicted by PhyloP27way conservation score averages for the *D. melanogaster* genome (black) and the Nup complex (grey). (**b**) Evolutionary conservation measured by PhyloP27way conservation score averages for the 350 nucleotide promoter regions compared between the Nup complex and the genome. Statistically significant differences from unpaired student t tests are indicated by asterisks (*** $p \le 0.0001$). (**c**) Evolutionary conservation measured by PhyloP27way conservation is shown as promoter d^P scores for the 350 nucleotide window in the promoters of the Nup complex and the genome averages. Statistically significant differences from non-parametric chi-squared tests are indicated by asterisks (** $p \le 0.001$). (**d–i**) Heatmaps indicating divergence for all changes (**d**), indels (**f**), and base changes (**h**) in yellow among closely related *D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba* and *D. erecta* for the control group of genes (m6A writer complex and readers) and the inner ring (IR), outer ring (OR), cytoplasmic filament (CF) and nuclear basket (NB) below the gene model with the TSS indicated by an arrow. The red line indicates the promoter region used for quantification of the substitution rate. Quantification of the change rate of all changes (**e**), indels (**g**), and base changes (**i**) in the TATA box distal region (from −30 to −380). Statistically significant differences from non-parametric chi-squared tests are indicated by asterisks (** $p \le 0.001$, ***$p \le 0.0001$).

Although the megadalton Nup complex has mainly been associated with providing a physical barrier between nucleus and cytoplasm, recent discovery of pleiotropic functions of channel Nup54 provide new insights how the Nup complex could drive speciation[15]. Nup54 is required for neuronal wiring underlying the female post-mating response induced by male-derived sex-peptide. Although we have not observed an obvious effect of incompatibility of Nup54 through sterility or lethality, this could be extended to be incompatibility resulting in incompatible behaviours. Considering the involvement of Nup54 in mating behaviour, one plausible scenario could involve exchange of sexual conflict for intragenomic conflict as the adaptive driver of initial divergence[10,13–15,18,43]. Here, the *Nup54* gene promoter could undergo rapid changes driven by sexual conflict. If this occurs in one or more isolated populations, pleiotropic effects of changes to Nup54 regulation of neuronal wiring, sexual differentiation or transposon regulation could result in genetic incompatibilities when the isolated population(s) come into contact with the original population[15].

Alternatively, Nups could drive speciation through their regulation of transposon silencing directly. Channel Nups are required for the generation of piRNAs originating from the *flamenco* locus in the germline to silence transposons[24]. Maternally inherited piRNAs are essential to transposon silencing and an imbalance can lead to a phenomenon called hybrid dysgenesis, imposing reduced fecundity in the female offspring when the male genome contributes novel transposable elements to be silenced[44]. Key to silencing are epigenetically inherited piRNAs from the female to prime the "ping-pong" cycle for amplification to increase the silencing capacity by heterochromatinization for preventing transposon mobilisation[44–47]. Compromised transposon silencing leading to deleterious reductions of fecundity could be escaped by altered Nup complex function. Such changes likely trigger pleiotropic effects that could include changes in neuronal wiring and behaviour. Such changes, however, could become fixed as a result of behavioural isolation, but restoring piRNA silencing regulation to its original state would not revert to the initial neurological state. In essence, if escape from lower fecundity as a result of hybrid dysgenesis[44–47], leading to reduced transposon silencing is coupled to changes in the morphology or behaviour and behavioural isolation, new species could be established very quickly. In particular, if changes in Nup functionality result in rewiring of neuronal circuits as indicated for Nup54, behavioural preferences leading to rapid isolation are likely. In fact, mate preference has changed in *D. simulans* through rewiring of sensory neuron projections to fruitless P1 neurons that control courtship[48].

Conclusion

Changes in gene expression have profound effects during species divergence and phenotypic adaptation, and such changes can lead to hybrids' gene mis-expression and dysfunction[49–53]. The molecular mechanism how new species arise through differential regulation of gene expression remains uncertain. The newly discovered role of channel Nups in piRNA processing in the germline to maintain transposon silencing provokes the claim that any compensation of the negative impact on fecundity from hybrid dysgenesis would be favoured. However, if as result of the pleiotropic effect of Nups changes in neuronal wiring and behaviour occur, these changes could be irreversible if behavioural isolation has already advanced[15,24]. Since mutations in promoters can directly affect gene expression regulation[54–56], such changes could be rapidly fixed in contrast to recessive changes in coding regions, which would remain hidden in heterozygosity. Our systematic analysis of the evolution of all Nups coding and promoter regions suggests a mode of evolution through changes of sequences upstream of the TSS, particularly in the promoters of core Nups, and supports the possibility that compensation of deleterious changes in the germline can lead to altered neuronal wiring and rapid fixation of adaptive traits.

Materials and methods

Open reading frame analysis. To determine whether the ORFs of nucleoporins are under selection, the PopFly online database (imkt.uab.cat) developed from the Drosophila Genome Nexus project assembling sequence data of around 1100 *D. melanogaster* genomes were used to perform MKTs to analyse polymorphism data from the ancestral Congo population because it is a sub-Saharan population with higher ancestral stability than other populations[57–59]. Synonymous and non-synonymous polymorphisms within *D. melanogaster* or between *D. melanogaster* and *D. simulans* were obtained and significance determined by Fisher's Exact Test, with significance defined as $p \leq 0.05$ after FDR correction. We used the concatenate and the compare against whole-genome distribution advanced options within the PopFly database analysis tool to compare the proportion of substitutions fixed by positive selection (α). We compared the 30 Nups against a random sample of the same size by randomly sampling from both datasets 100 times. Significance was calculated through an unpaired t test with significance defined as $p \leq 0.05$.

Sequence/data retrieval and alignment. Nucleoporin gene and promoter sequences for the five analysed *Drosophila* species (*D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba*, and *D. erecta*) were retrieved from UCSC genome browser (genome.ucsc.edu) using the Table Browser tool[37,38]. Pairwise and multiple alignments were carried out with clustalW using the alignment program version MEGA11 default alignment settings[60]. PhyloP27way data were sourced from UCSC genome browser through the Table Browser tool[37,38]. Data points were collected for a region of 1000 nucleotides upstream and 300 nucleotides downstream of gene TSSs relative to their transcription strands.

Accumulation of substitutions along extended gene regions. To test for nonrandom accumulation of indels along the Nup extended gene regions between the five analysed *Drosophila* species (*D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba*, and *D. erecta*) we determined significant deviations from a uniform distribution of substitutions using an empirical cumulative distribution function, as described by Civetta et al. (2016)[61]. The position of the indel event was defined as the 5' site of the start of the indel in the alignment[62]. The function

(G) detects monotonic increases in substitutions (n) measured as the difference between the relative occurrence of a nucleotide change and its relative position in the alignment[61]. Whether differences between the values of the G function (ΔG) between substitutional events deviates from a random accumulation of changes is tested using Monte Carlo simulations to produce 100,000 samples of n events by sampling sites without replacement along the alignment[61].

Comparison of substitution rates. To analyse the promoters of Nups and the control group of the m^6A writer complex and readers, a region of 1000 nucleotides upstream and downstream of the TSS was used with the TSS as an anchor point. Regions were aligned between the five analysed *Drosophila* species (*D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba*, and *D. erecta*) and alignments were translated into events[63]. Per Tang and Lewontin (1999)[63], 0 signified no sequence difference between all analysed species and 1 signified comparative sequence divergence for ≥ 1 species. The same was also performed for indels and base changes individually, where 0 signified no indel or base change event and 1 signified an indel or base change event for ≥ 1 species. A sliding window of five nucleotides upstream and downstream of each position was summed and processed to define sliding event (*Se*) scores which were used to generate heatmaps. *Se* scores were averaged between the different Nup subcomplexes and the m^6A control. To calculate sequence change accumulation (percentage of events greater than the average control promoter sliding window score (*d*)), a 350 nucleotide region upstream of the approximated TATA box region was taken. The total number of *Se* scores greater than the average control *Se* (Se^C) was divided by the total number of events in the region (*N*).

$$d = \frac{\text{Total number of Se events where } Se > Se^C}{N} \times 100$$

Significance was defined by non-parametric chi-squared tests versus the control score with one degree of freedom, where the percentage groupings were $> Se^C$ (*d*) and $\leq Se^C$. P values ≤ 0.05 with Bonferroni correction were deemed statistically significant.

To compare the conservation rate of the Nup complex compared to the *D. melanogaster* genome, PhyloP27way scores for a region of 1000 nucleotides upstream and 300 nucleotides downstream of the TSS for all *Drosophila* genes was performed. Genes without data points for the full 1300 nucleotide region were omitted. Genome-wide and Nup complex mean PhyloP scores were calculated using R version 4.4.2[64]. The average 350 nucleotide region upstream of the approximated TATA box region was compared between the two and significance was calculated using an unpaired student t-test. P values ≤ 0.05 were deemed statistically significant. For the same region of the Nup complex and genome average PhyloP *d* (d^P) was calculated, and significance determined as follows: the total number of PhyloP (*p*) scores in the 350 nucleotide regions less than the control group average promoter region (p^C) replaced the total number of *Se* events where $Se > Se^C$ (see below).

$$d^P = \frac{\text{Total number of } p \text{ scores } < p^C}{N} \times 100$$

Significance was defined by non-parametric chi-squared tests versus the control score with one degree of freedom, where the percentage groupings were $< p^C$ (d^P) and $\geq p^C$. P values ≤ 0.05 were deemed statistically significant.

Data availability
All data generated or analysed during this study are included in Supplementary Dataset 1.

References
1. Lin, D. H. & Hoelz, A. The structure of the nuclear pore complex (an update). *Annu. Rev. Biochem.* **88**, 725–783. https://doi.org/10.1146/annurev-biochem-062917-011901 (2019).
2. Cabal, G. G. *et al.* SAGA interacting factors confine sub-diffusion of transcribed genes to the nuclear envelope. *Nature* **441**, 770–773 (2006).
3. Pascual-Garcia, P. & Capelson, M. Nuclear pores as versatile platforms for gene regulation. *Curr. Opin. Genet. Dev.* **25**, 110–117 (2014).
4. Taddei, A. *et al.* Nuclear pore association confers optimal expression levels for an inducible yeast gene. *Nature* **441**, 774–778 (2006).
5. Breuer, M. & Ohkura, H. A negative loop within the nuclear pore complex controls global chromatin organization. *Genes Dev.* **29**, 1789–1794. https://doi.org/10.1101/gad.264341.115 (2015).
6. Avila, F. W., Sirot, L. K., LaFlamme, B. A., Rubinstein, C. D. & Wolfner, M. F. Insect seminal fluid proteins: Identification and function. *Annu. Rev. Entomol.* **56**, 21–40. https://doi.org/10.1146/annurev-ento-120709-144823 (2011).
7. Gillott, C. Male accessory gland secretions: Modulators of female reproductive physiology and behavior. *Annu. Rev. Entomol.* **48**, 163–184 (2003).
8. Kubli, E. Sex-peptides: seminal peptides of the Drosophila male. *Cell Mol. Life Sci.* **60**, 1689–1704 (2003).
9. Soller, M., Bownes, M. & Kubli, E. Control of oocyte maturation in sexually mature Drosophila females. *Dev. Biol.* **208**, 337–351 (1999).
10. Chapman, T. Sexual conflict: Mechanisms and emerging themes in resistance biology. *Am. Nat.* **192**, 217–229 (2018).
11. Coyne, J. A. & Orr, H. A. *Speciation* (Sinauer Associates, 2004).
12. Grayson, P. Izumo1 and Juno: The evolutionary origins and coevolution of essential sperm-egg binding partners. *R. Soc. Open Sci.* **2**, 150296. https://doi.org/10.1098/rsos.150296 (2015).
13. Hollis, B. *et al.* Sexual conflict drives male manipulation of female postmating responses in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **116**, 8437–8444. https://doi.org/10.1073/pnas.1821386116 (2019).

14. Hopkins, B. R. & Perry, J. C. The evolution of sex peptide: Sexual conflict, cooperation, and coevolution. *Biol. Rev.* https://doi.org/10.1111/brv.12849 (2022).
15. Nallasivan, M. P., Haussmann, I. U., Civetta, A. & Soller, M. Channel nuclear pore protein 54 directs sexual differentiation and neuronal wiring of female reproductive behaviors in Drosophila. *BMC Biol.* **19**, 1–14 (2021).
16. Patlar, B., Jayaswal, V., Ranz, J. M. & Civetta, A. Nonadaptive molecular evolution of seminal fluid proteins in Drosophila. *Evol. Int. J. Organ. Evolut.* **75**, 2102–2113 (2021).
17. Rowe, M. *et al.* Molecular diversification of the seminal fluid proteome in a recently diverged passerine species pair. *Mol. Biol. Evol.* **37**, 488–506. https://doi.org/10.1093/molbev/msz235 (2020).
18. Sirot, L. K., Wong, A., Chapman, T. & Wolfner, M. F. Sexual conflict and seminal fluid proteins: A dynamic landscape of sexual interactions. *Cold Spring Harb. Perspect Biol.* **7**, a017533. https://doi.org/10.1101/cshperspect.a017533 (2014).
19. Presgraves, D. C. The molecular evolutionary basis of species formation. *Nat. Rev. Genet.* **11**, 175–180. https://doi.org/10.1038/nrg2718 (2010).
20. Presgraves, D. C., Balagopalan, L., Abmayr, S. M. & Orr, H. A. Adaptive evolution drives divergence of a hybrid inviability gene between two species of Drosophila. *Nature* **423**, 715–719. https://doi.org/10.1038/nature01679 (2003).
21. Tang, S. & Presgraves, D. C. Evolution of the Drosophila nuclear pore complex results in multiple hybrid incompatibilities. *Science* **323**, 779–782 (2009).
22. Tang, S. & Presgraves, D. C. Lineage-specific evolution of the complex Nup160 hybrid incompatibility between *Drosophila melanogaster* and its sister species. *Genetics* **200**, 1245–1254 (2015).
23. Haussmann, I. U., Hemani, Y., Wijesekera, T., Dauwalder, B. & Soller, M. Multiple pathways mediate the sex-peptide-regulated switch in female Drosophila reproductive behaviours. *Proc. Biol. Sci.* **280**, 20131938. https://doi.org/10.1098/rspb.2013.1938 (2013).
24. Munafo, M. *et al.* Channel nuclear pore complex subunits are required for transposon silencing in Drosophila. *Elife* **10**, e66321. https://doi.org/10.7554/eLife.66321 (2021).
25. Smith, N. G. C. & Eyre-Walker, A. Adaptive protein evolution in Drosophila. *Nature* **415**, 1022–1024. https://doi.org/10.1038/4151022a (2002).
26. Croze, M. *et al.* A genome-wide scan for genes under balancing selection in Drosophila melanogaster. *BMC Evolut. Biol.* https://doi.org/10.1186/s12862-016-0857-z (2017).
27. Balacco, D. L. & Soller, M. The m(6)A writer: Rise of a machine for growing tasks. *Biochemistry* **58**, 363–378. https://doi.org/10.1021/acs.biochem.8b01166 (2019).
28. Bawankar, P. *et al.* Hakai is required for stabilization of core components of the m(6)A mRNA methylation machinery. *Nat. Commun.* **12**, 3778. https://doi.org/10.1038/s41467-021-23892-5 (2021).
29. Dezi, V., Ivanov, C., Haussmann, I. U. & Soller, M. Nucleotide modifications in messenger RNA and their role in development and disease. *Biochem. Soc. Trans.* **44**, 1385–1393. https://doi.org/10.1042/bst20160110 (2016).
30. Knuckles, P. *et al.* Zc3h13/Flacc is required for adenosine methylation by bridging the mRNA-binding factor Rbm15/Spenito to the m(6)A machinery component Wtap/Fl(2)d. *Genes Dev.* **32**, 415–429. https://doi.org/10.1101/gad.309146.117 (2018).
31. Su, S. *et al.* Cryo-EM structures of human m6A writer complexes. *Cell Res.* **32**, 982–994. https://doi.org/10.1038/s41422-022-00725-8 (2022).
32. Hoogendoorn, B. *et al.* Functional analysis of human promoter polymorphisms. *Hum. Mol. Genet.* **12**, 2249–2254. https://doi.org/10.1093/hmg/ddg246 (2003).
33. Smale, S. T. & Kadonaga, J. T. The RNA polymerase II core promoter. *Annu. Rev. Biochem.* **72**, 449–479. https://doi.org/10.1146/annurev.biochem.72.121801.161520 (2003).
34. Vo Ngoc, L., Kassavetis, G. A. & Kadonaga, J. T. The RNA polymerase II core promoter in Drosophila. *Genetics* **212**, 13–24. https://doi.org/10.1534/genetics.119.302021 (2019).
35. Hill, M. S., Vande Zande, P. & Wittkopp, P. J. Molecular and evolutionary processes generating variation in gene expression. *Nat. Rev. Genet.* **22**, 203–215. https://doi.org/10.1038/s41576-020-00304-w (2021).
36. Hoskins, R. A. *et al.* Genome-wide analysis of promoter architecture in *Drosophila melanogaster*. *Genome Res.* **21**, 182–192. https://doi.org/10.1101/gr.112466.110 (2011).
37. Karolchik, D. The UCSC table browser data retrieval tool. *Nucleic Acids Res.* **32**, 493D – 496. https://doi.org/10.1093/nar/gkh103 (2004).
38. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006. https://doi.org/10.1101/gr.229102 (2002).
39. Consortium D.G. Evolution of genes and genomes on the Drosophila phylogeny. *Nature* **450**, 203–218. https://doi.org/10.1038/nature06341 (2007).
40. Main, B. J., Smith, A. D., Jang, H. & Nuzhdin, S. V. Transcription start site evolution in Drosophila. *Mol. Biol. Evol.* **30**, 1966–1974. https://doi.org/10.1093/molbev/mst085 (2013).
41. Anreiter, I., Tian, Y. W. & Soller, M. The cap epitranscriptome: Early directions to a complex life as mRNA. *BioEssays* https://doi.org/10.1002/bies.202200198 (2022).
42. Taylor, M. S. *et al.* Heterotachy in mammalian promoter evolution. *PLoS Genet.* **2**, e30 (2006).
43. Gardner, A. & Úbeda, F. The meaning of intragenomic conflict. *Nat. Ecol. Evolut.* **1**, 1807–1815. https://doi.org/10.1038/s41559-017-0354-9 (2017).
44. Czech, B. *et al.* piRNA-guided genome defense: From biogenesis to silencing. *Annu. Rev. Genet.* **52**, 131–157. https://doi.org/10.1146/annurev-genet-120417-031441 (2018).
45. Blumenstiel, J. P. & Hartl, D. L. Evidence for maternally transmitted small interfering RNA in the repression of transposition in *Drosophila virilis*. *Proc. Natl. Acad. Sci. USA* **102**, 15965–15970. https://doi.org/10.1073/pnas.0508192102 (2005).
46. Malone, C. D. & Hannon, G. J. Molecular evolution of piRNA and transposon control pathways in Drosophila. *Cold Spring Harb. Symp. Quant. Biol.* **74**, 225–234. https://doi.org/10.1101/sqb.2009.74.052 (2009).
47. Brennecke, J. *et al.* An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* **322**, 1387–1392. https://doi.org/10.1126/science.1165171 (2008).
48. Seeholzer, L. F., Seppo, M., Stern, D. L. & Ruta, V. Evolution of a central neural circuit underlies Drosophila mate preferences. *Nature* **559**, 564–569. https://doi.org/10.1038/s41586-018-0322-9 (2018).
49. Cutter, A. D. & Bundus, J. D. Speciation and the developmental alarm clock. *Elife* https://doi.org/10.7554/eLife.56276 (2020).
50. Hagen, J. F. D. *et al.* Tartan underlies the evolution of Drosophila male genital morphology. *Proc. Natl. Acad. Sci. USA* **116**, 19025–19030. https://doi.org/10.1073/pnas.1909829116 (2019).
51. Mack, K. L. & Nachman, M. W. Gene regulation and speciation. *Trends Genet.* **33**, 68–80. https://doi.org/10.1016/j.tig.2016.11.003 (2017).
52. Civetta, A. Misregulation of gene expression and sterility in interspecies hybrids: Causal links and alternative hypotheses. *J. Mol. Evol.* **82**, 176–182. https://doi.org/10.1007/s00239-016-9734-z (2016).
53. King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116. https://doi.org/10.1126/science.1090005 (1975).
54. Wittkopp, P. J., Haerum, B. K. & Clark, A. G. Evolutionary changes in cis and trans gene regulation. *Nature* **430**, 85–88 (2004).
55. Pastinen, T. *et al.* A survey of genetic and epigenetic variation affecting human gene expression. *Physiol. Genomics* **16**, 184–193 (2004).
56. Ronald, J., Brem, R. B., Whittle, J. & Kruglyak, L. Local regulatory variation in *Saccharomyces cerevisiae*. *PLoS Genet.* **1**, e25 (2005).

57. Lack, J. B. *et al.* The Drosophila genome nexus: A population genomic resource of 623 *Drosophila melanogaster* genomes, including 197 from a single ancestral range population. *Genetics* **199**, 1229–1241. https://doi.org/10.1534/genetics.115.174664 (2015).
58. Lack, J. B., Lange, J. D., Tang, A. D., Corbett-Detig, R. B. & Pool, J. E. A thousand fly genomes: An expanded Drosophila genome nexus. *Mol. Biol. Evol.* **33**, 3308–3313. https://doi.org/10.1093/molbev/msw195 (2016).
59. Mackay, T. F. *et al.* The *Drosophila melanogaster* genetic reference panel. *Nature* **482**, 173–178. https://doi.org/10.1038/nature10811 (2012).
60. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549. https://doi.org/10.1093/molbev/msy096 (2018).
61. Civetta, A., Ostapchuk, D. C. & Nwali, B. Genome hotspots for nucleotide substitutions and the evolution of Influenza A (H1N1) human strains. *Genome Biol. Evol.* **8**, 986–993. https://doi.org/10.1093/gbe/evw061 (2016).
62. Schaeffer, S. W. Molecular population genetics of sequence length diversity in the *Adh* region of *Drosophila pseudoobscura*. *Genet. Res.* **80**, 163–175. https://doi.org/10.1017/s0016672302005955 (2002).
63. Tang, H. & Lewontin, R. C. Locating regions of differential variability in DNA and protein sequences. *Genetics* **153**, 485–495. https://doi.org/10.1093/genetics/153.1.485 (1999).
64. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, 2022).

Acknowledgements

Author contributions

M.S. conceived and directed the project. All authors analysed data. M.S., D.W.J.M. and A.C. wrote the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-34985-0.

Correspondence and requests for materials should be addressed to A.C. or M.S.

Reprints and permissions information is available at www.nature.com/reprints.

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