Indel driven rapid evolution of core nuclear pore protein gene promoters

DAVID W. J. MCQUARRIE[1,2*], ADAM M. READ[1*], FRANNIE H. S. STEPHENS[1], ALBERTO CIVETTA[3,4] AND MATTHIAS SOLLER[1,2,4]

[1]School of Biosciences, College of Life and Environmental Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom

[2]Birmingham Centre for Genome Biology, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom

[3]Department of Biology, University of Winnipeg, Winnipeg, Canada MB R3B 2E9

[4]Corresponding authors: m.soller@bham.ac.uk and a.civetta@uwinnipeg.ca

*equal contributing authors

Running title: Evolution of core nucleoporin gene promoters

Key Words: Nuclear pore complex (NPC), NPC ring, rapid evolution, promoter evolution, speciation, piRNA expression, neuronal wiring, transposon silencing

Abstract

Nuclear pore proteins (Nups) prominently are among the few genes linked to speciation from hybrid incompatibility in *Drosophila*. It was previously found that neuronal wiring underlying the female post-mating response induced by male-derived sex-peptide requires channel Nup54 functionality. A hot spot for rapid evolution in the promoter of *Nup54* suggests a critical role for regulatory elements at the onset of speciation. Systematic analysis of Nup coding and promoter regions using *Drosophila* phylogenomics reveals that polymorphism differences between closely related *Drosophila* species in Nup coding regions do not generally evolve rapidly. Consistent with findings for *Nup54*, additional channel Nups 58 and 62 promotors are also hotpots for rapid accumulation of insertions/deletions (indels). Examination of Nup upstream regions reveals that core nuclear pore complex gene promoters accumulate indels rapidly. Since changes in promoters can have dominant effects (effects which directly impact gene expression of associated genes), these results indicate an evolutionary mechanism driven by indel accumulation in core Nup promoters. Compensation of such deleterious changes could lead to altered neuronal wiring, rapid fixation of adaptive traits 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.

Introduction

Male derived substances transferred during mating change female physiology and behaviour to guarantee reproductive success in many insects [1-4]. 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 [5-13]. Such hybrid incompatibility among closely related *Drosophila* species has been used to identify responsible genes. Among the few genes identified were nuclear pore proteins

Nup98 and Nup160 [14-17]. In addition, a screen for sex-peptide insensitive mutants identified Nup54 functionality to be important for neuronal wiring of circuits involved in regulating post-mating behaviours [10]. 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 [10]. Moreover, this *Nup54* promoter deletion allele maps to a hot spot for rapid evolution in the *Nup54* promotor suggesting sexual conflict driving female escape from male manipulation by sex-peptide under unfavourable conditions [10]. Furthermore, channel Nups have also been attributed a role in 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 [18], indicating that the disruption to NPC regulation likely has pleiotropic effects.

The nuclear pore protein complex (NPC) provides a physical barrier between nucleus and cytoplasm requiring active transport for cargos above about 40 kDa [19]. The NPC consists of 30 evolutionary highly conserved proteins 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 NPC can contribute to differential expression of other genes as e.g. actively transcribed genes can be in the proximity of the NPC [20-23].

Given the rapid evolution of *Nup54* gene expression control and the links of Nups to speciation, we systematically examined, whether the NPC is under increased evolutionary constraint as the molecular mechanisms of how the NPC is involved in speciation remains unexplained. Our systematic analysis of Nup coding and promoter regions using *Drosophila* phylogenomics reveals promoters of Nup core genes as hot spots for insertion/deletion (indel)

driven rapid evolution and the nuclear pore complex as a nexus for species-specific changes *via* nucleo-cytoplasmic transport regulated gene expression.

Results

Rapid evolution of core nuclear pore protein genes is not generally driven through adaptation in coding sequences

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 between *D. melanogaster* and its closest relative *D. simulans*. This analysis showed that most Nup ORFs do not deviate from neutral evolution (Fig. 1b), but that 30% (3 out of 10) of the Nups constituting the IR showed significant evidence of selection.

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 hot spot for rapid evolution [10], 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 analysis revealed that also the promoters of the other two channel Nups are hot spots for rapid evolution (Fig. 2). Moreover, these changes contain a number of indels (Fig. 2c and d) that can fundamentally impact on transcription factor binding commonly found in the proximal region before the TATA box [24-28].

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, Supplementary Fig. S1). We analysed conservation of in 27 insect species through the PhyloP27way data [29,30]. To determine whether the NPC evolves rapidly compared to the genome, we used PhyloP scores as a measure of conservation upstream of the TSS (Fig. 3a). We observed a significant decrease in NPC sequence conservation in the promoter region upstream of the predicted TATA box site (-30 to -380) (Fig. 3b and c). To focus more specifically on *Drosophila*, we analysed promoter evolution between species in the *melanogaster* subgroup [31]. As a control group we used the m^6A mRNA methylation machinery (consisting of Mettl3, Mettl14, fl(2)d, virilizer, flacc, nito and Hakai and the two 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 promoter evolution [32-35]. In contrast to the genes from the m^6A mRNA methylation machinery, promoters of ribosomal genes are different in their transcription start site (TSS) properties [36,37].

As has been observed for *Drosophila* and human promoters [28,36,38], 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 m^6A mRNA methylation writer complex and readers consistent with the general trend (Fig. 3d and e, Supplementary Fig. 1). We then analysed the promoters of the different Nup sub-complexes IR, OR, CF and NF. 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. 3 and Supplementary Fig. S1c-f). We next analysed all sequence differences and performed a more detailed analysis split into deletions, insertions and base changes, relative to *D. melanogaster*. Analysis of all sequence differences reflected our indel

analysis, with a significant increase in the number of differences in IR and OR Nups (Supplementary Fig. S1a and b). Our more detailed analysis revealed a significant increase in the number of deletions in IR, OR and to a lesser extend in the cytoplasmic filaments (CF) compared to the control (Supplementary Fig. S1c and d), while no significant changes were detected for insertions (Supplementary Fig. S1e and f). In contrast base changes are increased for IR and reduced for OR and CF (Supplementary Fig. S1g and h).

Discussion

Here, through comparison of polymorphism rates between closely related Drosophila species we show that rapid evolution of Nups is not generally driven through changes in ORFs. Focussing on upstream and downstream Nup gene regions, we reveal a bias for inclusion of nucleotide changes in the core members of the NPC. Consistent with channel *Nup54* [10], further analysis of upstream regions of the other channel Nups 58 and 62 reveals their promotor regions to be hotspots for rapid indel accumulation. Through examination of promoter regions of Nup subcomplex members, we identify the core IR and OR Nups as hot spots for rapid accumulation of indels. Here, the core NPC members likely act as drivers of species-specific variation through dominant indel driven changes in promoters, subsequently modifying regulation of nucleo-cytoplasmic transport regulated gene expression.

Although the megadalton nuclear pore protein complex (NPC) 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 NPC could drive speciation. Nup54 is required for neuronal wiring underlying the female post-mating response induced by male-derived sex-peptide. Since its promoter is a hot spot for rapid evolution, a critical role for regulatory elements at the onset of speciation is indicated as a result of sexual conflict [5,9,10,13,39].

Given the role of channel Nups in piRNA meditated transposon silencing of the *flamenco* locus in the germline and their requirement for sexual differentiation [10,18], changes at one trait could cascade into the second leading to rapid phenotypic evolution and speciation. Intriguingly, mutations in *cis*-regulatory sequences of genes are often dominant whereas the majority of coding mutations are recessive [40-42] leading to different efficacy of selection [43-45].

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 [46]. 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 [46-49]. Compromised transposon silencing leading to deleterious reductions of fecundity could be escaped by altered NPC function. Such changes are likely to trigger pleiotropic effects including altered neuronal wiring and change of behaviour, leading to accelerated evolution and speciation. In essence, if escape from lower fecundity as a result of hybrid dysgenesis [46-49], leading to reduced transposon silencing is coupled to changes in the morphology or behaviour, 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 could lead to rapid isolation. In fact, mate preference has changed in *D. simulans* through rewiring of sensory neuron projections to fruitless P1 neurons that control courtship [50].

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 [51-55]. 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, due to the pleiotropic effect of Nups on neuronal wiring and behaviour, the escape from hybrid incompatibility could produce changes leading to behavioural isolation [10,18]. Since mutations in promoters can exert a dominant effect [40-42], 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 less studied mode of evolution through changes of sequences upstream of the transcription start site (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 [56-58]. 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 $p \leq 0.05$ with Bonferroni correction.

Sequence/data retrieval and alignment

Nucleoporin gene and promoter sequences were retrieved from UCSC genome browser (genome.ucsc.edu) using the Table Browser tool [29,30]. Alignments were carried out with clustalW within the MEGA11 program [59]. PhyloP27way data were sourced from UCSC genome browser through the Table Browser tool [29,30]. Data points were collected for a region of 1000 nucleotides upstream and 300 nucleotides downstream.

Accumulation of substitutions along extended gene regions

To test for nonrandom accumulation of substitutions along the Nup extended gene regions we determined significant deviations from a uniform distribution of substitutions using an empirical cumulative distribution function. 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 [60]. 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 [60].

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 bp upstream and downstream of the transcriptional start site (TSS) was used with TSS as an anchor point. Regions of differential alignment between the *Drosophila* species were averaged across the NPC subcomplexes, where alignments were translated into events [61]. Per [61], 0 signified no sequence difference between all analysed species and 1 signified comparative sequence divergence 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. 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 control *Se* (Se^C) was divided by the total number of events in the region (*N*).

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

Significance was defined by non-parametric chi-squared tests vs the control score. P values ≤0.05 were deemed statistically significant.

To compare the conservation rate of the NPC 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 NPC mean PhyloP scores were calculated using R version 4.4.2 [62]. 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 NPC and genome averages *d* was calculated and significance determined as described.

Data availability

All data generated or analysed during this study are included in Supplementary Dataset S1.

Acknowledgements

This work was supported by the Medical Research Council Integrated Midlands Partnership for Biomedical Training Program by D.W.J.M., the Biotechnology and Biological Science Research Council to M.S. and the Natural Sciences and Engineering Research Council of Canada to A.C. We thank R. Arnold for help with bioinformatics.

Authors 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 includes Fig. S1 and Supplementary Dataset 1.

References

1 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, (2011).

2 Gillott, C. Male accessory glandd secretions: Modulators of Female Reproductive Physiology and Behavior. *Annu. Rev. Entomol.* **48**, 163-184 (2003).

3 Kubli, E. Sex-peptides: seminal peptides of the Drosophila male. *Cell Mol Life Sci* **60**, 1689-1704 (2003).

4 Soller, M., Bownes, M. & Kubli, E. Control of oocyte maturation in sexually mature Drosophila females. *Dev Biol* **208**, 337-351 (1999).

5 Chapman, T. Sexual conflict: mechanisms and emerging themes in resistance biology. *The American Naturalist* **192**, 217-229 (2018).

6 Coyne, J. A. & Orr, H. A. *Speciation*. (Sinauer Associates, 2004).

7 Grayson, P. Izumo1 and Juno: the evolutionary origins and coevolution of essential sperm-egg binding partners. *Royal Society Open Science* **2**, 150296, (2015).

8 Hollis, B. *et al.* Sexual conflict drives male manipulation of female postmating responses in Drosophila melanogaster. *Proc Natl Acad Sci U S A* **116**, 8437-8444, (2019).

9 Hopkins, B. R. & Perry, J. C. The evolution of sex peptide: sexual conflict, cooperation, and coevolution. *Biological Reviews*, (2022).

10 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 biology* **19**, 1-14 (2021).

11 Patlar, B., Jayaswal, V., Ranz, J. M. & Civetta, A. Nonadaptive molecular evolution of seminal fluid proteins in Drosophila. *Evolution; International Journal of Organic Evolution* **75**, 2102 - 2113 (2021).

12 Rowe, M. *et al.* Molecular Diversification of the Seminal Fluid Proteome in a Recently Diverged Passerine Species Pair. *Mol Biol Evol* **37**, 488-506, (2020).

13 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, (2014).

14 Presgraves, D. C. The molecular evolutionary basis of species formation. *Nature Reviews Genetics* **11**, 175-180 (2010).

15 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 (2003).

16 Tang, S. & Presgraves, D. C. Evolution of the Drosophila nuclear pore complex results in multiple hybrid incompatibilities. *Science* **323**, 779-782 (2009).

17 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).

18 Munafo, M. *et al.* Channel nuclear pore complex subunits are required for transposon silencing in Drosophila. *Elife* **10**, e66321, (2021).

19 Lin, D. H. & Hoelz, A. The Structure of the Nuclear Pore Complex (An Update). *Annu Rev Biochem* **88**, 18.11-18.59, (2019).

20 Cabal, G. G. *et al.* SAGA interacting factors confine sub-diffusion of transcribed genes to the nuclear envelope. *Nature* **441**, 770-773 (2006).

21 Pascual-Garcia, P. & Capelson, M. Nuclear pores as versatile platforms for gene regulation. *Current opinion in genetics & development* **25**, 110-117 (2014).

22 Taddei, A. *et al.* Nuclear pore association confers optimal expression levels for an inducible yeast gene. *Nature* **441**, 774-778 (2006).

23 Breuer, M. & Ohkura, H. A negative loop within the nuclear pore complex controls global chromatin organization. *Genes & Development* **29**, 1789-1794, (2015).

24 Hoogendoorn, B. *et al.* Functional analysis of human promoter polymorphisms. *Human Molecular Genetics* **12**, 2249-2254, (2003).

25 Smale, S. T. & Kadonaga, J. T. The RNA polymerase II core promoter. *Annu Rev Biochem* **72**, 449-479, (2003).

26 Vo Ngoc, L., Kassavetis, G. A. & Kadonaga, J. T. The RNA Polymerase II Core Promoter in Drosophila. *Genetics* **212**, 13-24, (2019).

27 Hill, M. S., Vande Zande, P. & Wittkopp, P. J. Molecular and evolutionary processes generating variation in gene expression. *Nature Reviews Genetics* **22**, 203-215, (2021).

28 Hoskins, R. A. *et al.* Genome-wide analysis of promoter architecture in Drosophila melanogaster. *Genome Research* **21**, 182-192, (2011).

29 Karolchik, D. The UCSC Table Browser data retrieval tool. *Nucleic Acids Research* **32**, 493D-496, (2004).

30 Kent, W. J. *et al.* The Human Genome Browser at UCSC. *Genome Research* **12**, 996-1006, (2002).

31 Consortium, D. G. Evolution of genes and genomes on the Drosophila phylogeny. *Nature* **450**, 203-218, (2007).

32 Balacco, D. L. & Soller, M. The m(6)A Writer: Rise of a Machine for Growing Tasks. *Biochemistry* **58**, 363-378, (2019).

33 Bawankar, P. *et al.* Hakai is required for stabilization of core components of the m(6)A mRNA methylation machinery. *Nat Commun* **12**, 3778, (2021).

34 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, (2016).

35 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, (2018).

36 Main, B. J., Smith, A. D., Jang, H. & Nuzhdin, S. V. Transcription Start Site Evolution in Drosophila. *Molecular Biology and Evolution* **30**, 1966-1974, (2013).

37 Anreiter, I., Tian, Y. W. & Soller, M. The cap epitranscriptome: Early directions to a complex life as mRNA. *BioEssays*, 2200198, (2022).

38 Taylor, M. S. *et al.* Heterotachy in mammalian promoter evolution. *PLoS genetics* **2**, e30 (2006).

39 Hollis, B. *et al.* Sexual conflict drives male manipulation of female postmating responses in Drosophila melanogaster. *Proceedings of the National Academy of Sciences* **116**, 8437-8444 (2019).

40 Wittkopp, P. J., Haerum, B. K. & Clark, A. G. Evolutionary changes in cis and trans gene regulation. *Nature* **430**, 85-88 (2004).

41 Pastinen, T. *et al.* A survey of genetic and epigenetic variation affecting human gene expression. *Physiological genomics* (2004).

42 Ronald, J., Brem, R. B., Whittle, J. & Kruglyak, L. Local regulatory variation in Saccharomyces cerevisiae. *PLoS genetics* **1**, e25 (2005).

43 Wilkins, A. S. *The evolution of developmental pathways*. (Sinauer Associates, 2002).

44 Stern, D. L. Perspective: evolutionary developmental biology and the problem of variation. *Evolution* **54**, 1079-1091 (2000).

45 Wray, G. A. *et al.* The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol* **20**, 1377-1419, (2003).

46 Czech, B. *et al.* piRNA-Guided Genome Defense: From Biogenesis to Silencing. *Annu Rev Genet* **52**, 131-157, (2018).

47 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 U S A* **102**, 15965-15970, (2005).

48 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, (2009).

49 Brennecke, J. *et al.* An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* **322**, 1387-1392, (2008).

50 Seeholzer, L. F., Seppo, M., Stern, D. L. & Ruta, V. Evolution of a central neural circuit underlies Drosophila mate preferences. *Nature* **559**, 564-569, (2018).

51 Cutter, A. D. & Bundus, J. D. Speciation and the developmental alarm clock. *Elife* **9**, (2020).

52 Hagen, J. F. D. *et al.* tartan underlies the evolution of Drosophila male genital morphology. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 19025-19030, (2019).

53 Mack, K. L. & Nachman, M. W. Gene Regulation and Speciation. *Trends Genet* **33**, 68-80, (2017).

54 Civetta, A. Misregulation of Gene Expression and Sterility in Interspecies Hybrids: Causal Links and Alternative Hypotheses. *J Mol Evol* **82**, 176-182, (2016).

55 King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107-116, (1975).

56 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, (2015).

57 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, (2016).

58 Mackay, T. F. *et al.* The Drosophila melanogaster Genetic Reference Panel. *Nature* **482**, 173-178, (2012).

59 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, (2018).

60 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, (2016).

61 Tang, H. & Lewontin, R. C. Locating Regions of Differential Variability in DNA and Protein Sequences. *Genetics* **153**, 485-495, (1999).

62 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (2022).

Figure legends

Figure 1: Core nuclear pore protein genes do not generally evolve through adaptation in coding sequences (ORFs).

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 (yellow) and pore membrane proteins (POMS, red) embedded in the nuclear envelope lipid-bilayer (green).

b) Significance of synonymous to non-synonymous changes in open reading frames (ORFs) indicated by an asterisk and p-values.

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* 3′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 NPC (grey).

b) Evolutionary conservation measured by PhyloP27way conservation score averages for the 350-nucleotide promoter regions compared between the NPC and the genome. Statistically significant differences from unpaired student t-tests are indicated by asterisks (* $p \leq 0.05$, ** $p \leq 0.001$, *** $p \leq 0.0001$).

c) Evolutionary conservation measured by PhyloP27way conservation promoter *d* scores for the 350-nucleotide for the promoters of the NPC and the genome averages. Statistically significant differences from non-parametric chi-squared tests are indicated by asterisks (* $p \leq 0.05$, ** $p \leq 0.001$, *** $p \leq 0.0001$).

d) Heatmaps indicating divergence for indels in yellow among closely related *D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba* and *D. erecta* for the control group of genes (m^6A 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 transcription start site (TSS) indicated by an arrow. The red line indicates the promoter region used for quantification of the substitution rate.

e) Quantification of the change rate of indels in the TATA box distal region (from -30 to -380). Statistically significant differences from non-parametric chi-squared tests are indicated by asterisks (* $p \leq 0.05$, ** $p \leq 0.001$, *** $p \leq 0.0001$).

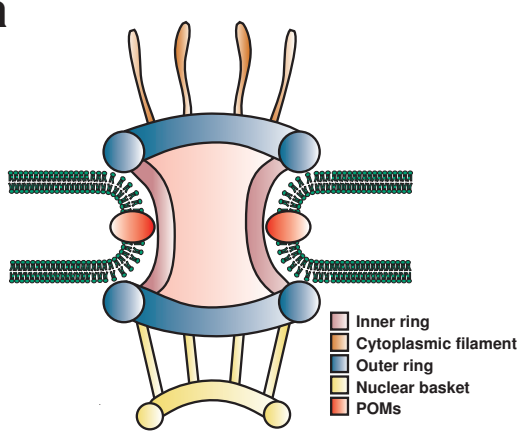
Supplementary Figure S1: Inner and outer ring nuclear pore protein gene subcomplexes undergo rapid evolution through indel accumulation in promoters.

a) Gene model depicting the intergenic region followed by the transcription start site (TSS) and the genespan with the 350-nucleotide region analysed for promotor activity upstream of the TATA box location.

a-h) Heatmaps indicating divergence in yellow among closely related Drosophila species 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). Calculated sequence difference index, deletion index, insertion index and base change index events are reported as bar charts for the NPC subgroups compared to the control group. Index events were defined by values above the control intergenic region average. Statistically significant differences between the NPC subgroups and the control group were calculated from chi-squared tests and indicated by asterisks (* $p \leq 0.05$, ** $p \leq 0.001$, *** $p \leq 0.0001$).

Figure 1

a



b

	Gene ID	ORF change significance
Inner Ring	*Nup205*	* 9×10^{-14}
	Nup188	* 3×10^{-08}
	Nup93-1	0.002
	Nup93-2	0.753
	Nup154	* 5×10^{-05}
	Nup35	0.010
	Nup54	0.685
	Nup58	1.000
	Nup62	0.402
	Nup98	0.033
Cytoplasmic filament	*Rae1*	0.003
	CG18787	1.000
	Mbo	1.000
	Nup214	0.326
	Gle1	0.703
	Nup358	0.235

	Gene ID	ORF change significance
Outer Ring	*Nup160*	0.002
	Nup75	0.016
	Nup44a	0.184
	Nup96	0.033
	Sec13	0.638
	Nup107	1.000
	Nup133	0.209
	Nup43	0.077
	Nup37	1.000
	ELYS	0.164
Nuclear basket	*Nup153*	0.051
	Nup50	0.599
	Mtor	0.007
POM	*NDC1*	0.592
	POM210	0.172

Figure 2

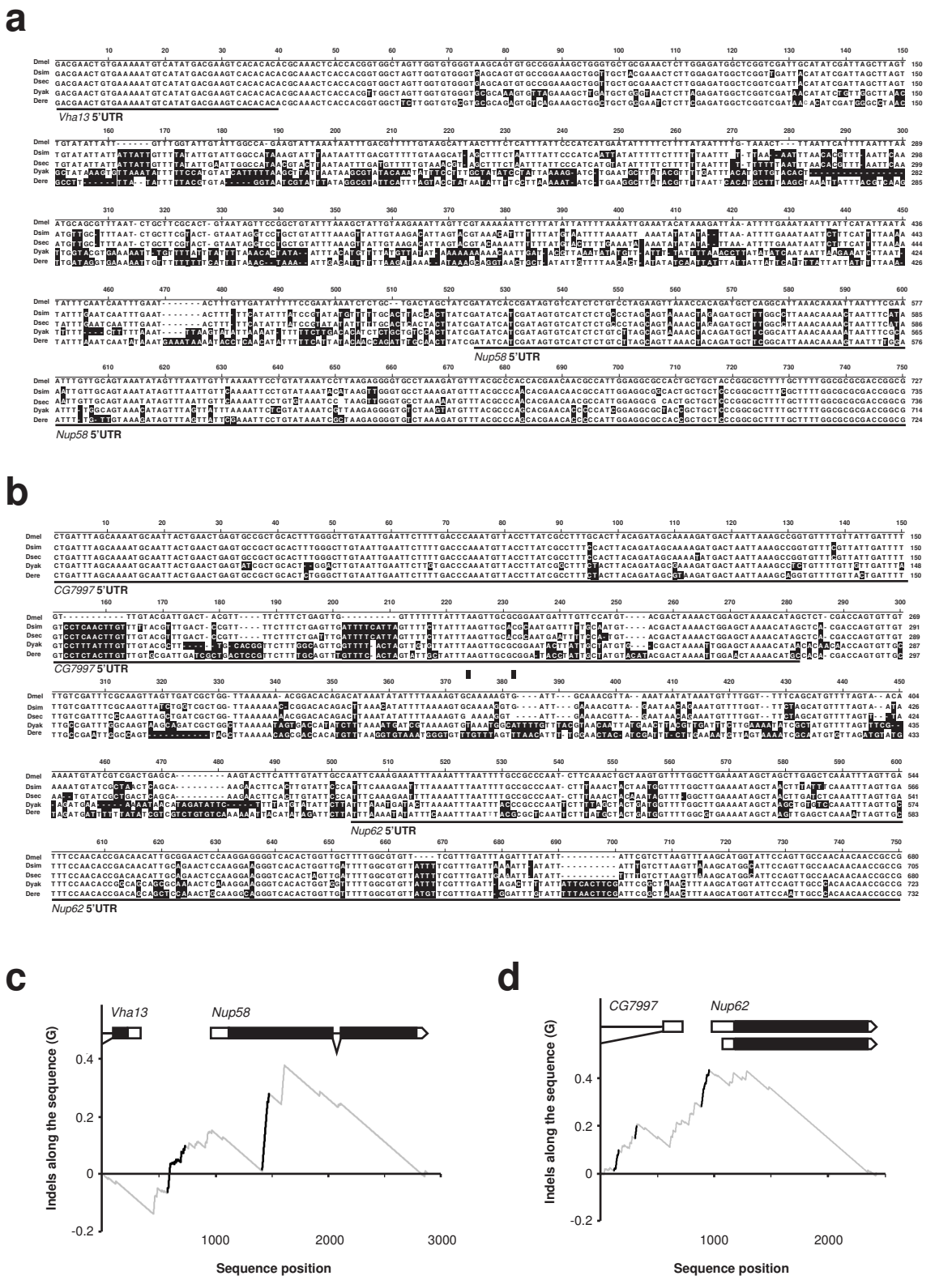
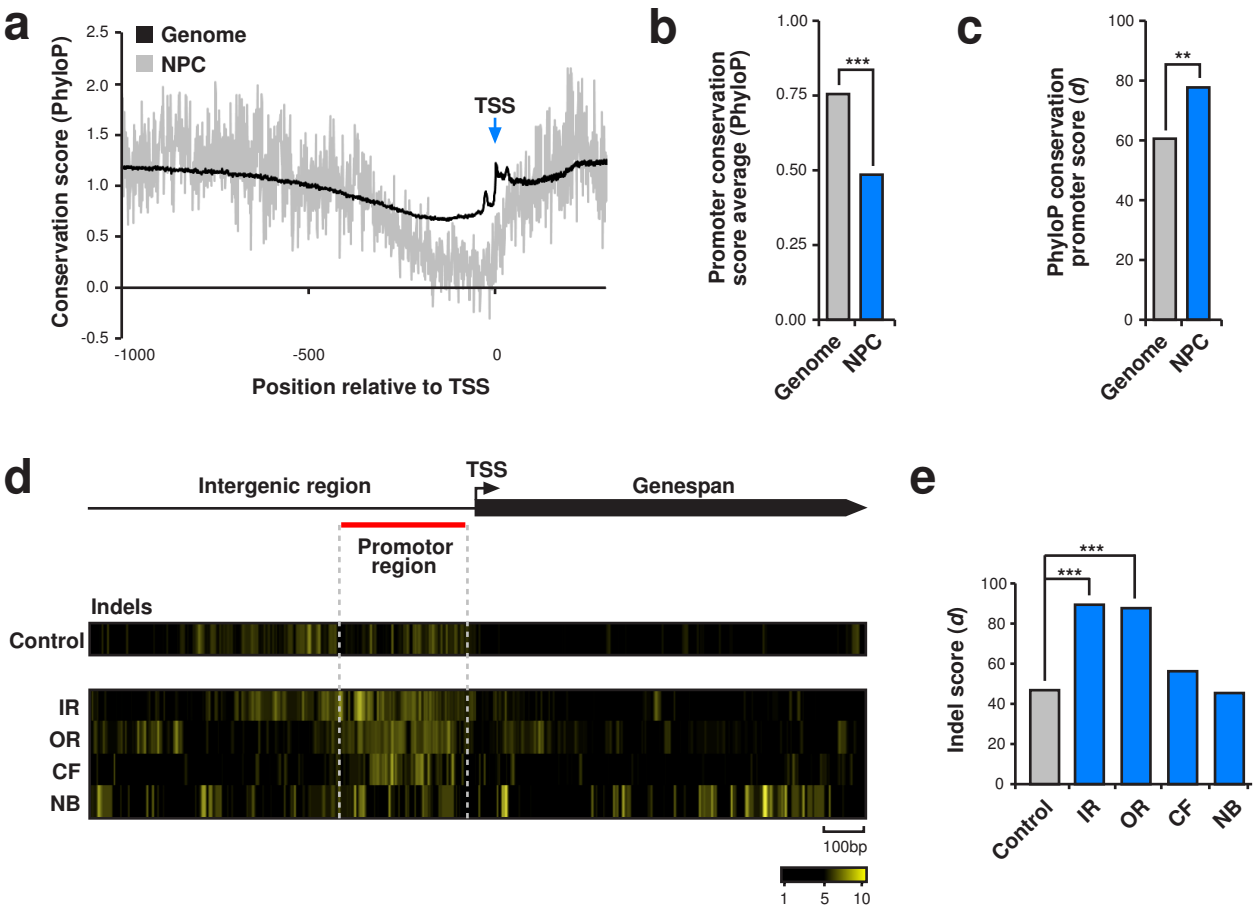
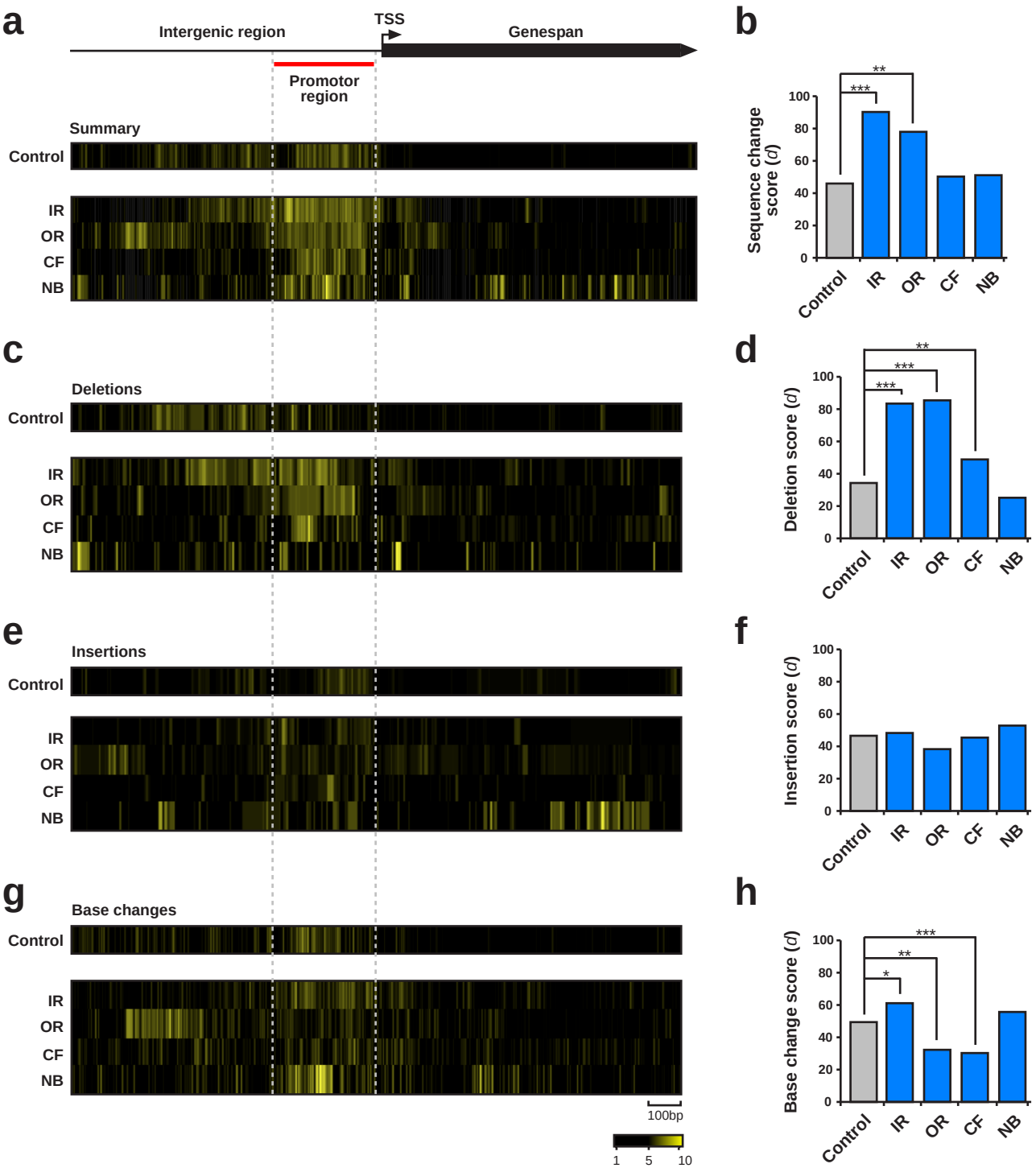

a
Dmel
Dsim
Dsec
Dyak
Dere
Vha13 5'UTR
Nup58 5'UTR
b
CG7997 5'UTR
Nup62 5'UTR
c
Vha13
Nup58
Indels along the sequence (G)
Sequence position
d
CG7997
Nup62
Indels along the sequence (G)
Sequence position

Figure 3


a
Genome
NPC
TSS
Conservation score (PhyloP)
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-1000
-500
0
Position relative to TSS
b
Promoter conservation score average (PhyloP)
1.00
0.75
0.50
0.25
0.00
***
Genome
NPC
c
PhyloP conservation promoter score (d)
100
80
60
40
20
0
**
Genome
NPC
d
Intergenic region
TSS
Genespan
Promotor region
Indels
Control
IR
OR
CF
NB
100bp
1
5
10
e
Indel score (d)
100
80
60
40
20
0
***
***
Control
IR
OR
CF
NB

Supplementary Figure S1