Indel driven rapid evolution of core nuclear pore protein gene promoters

DAVID W. J. MCQUARRIE[1,2*], ADAM M. READ[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*. Moreover, 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 open reading frames do not generally drive rapid evolution. Analysis of Nup gene upstream regions indicated a bias for changes compared to those downstream. Consistent with findings for *Nup54*, additional channel Nups 58 and 62 promotor regions are also hotpots for rapid accumulation of insertions/deletions (indels). Further examination of upstream regions of all Nup genes reveals that core nuclear pore complex gene promoters accumulate indels at a faster rate than the control. Since changes in promoters can have dominant effects (an effect that directly impacts gene expression of the associated gene), these results indicate a mechanism of evolution driven through accumulation of indels in core Nup gene 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

47 hybrids [5-13]. Such hybrid incompatibility among closely related *Drosophila* species has been
48 used to identify responsible genes. Among the few genes identified were nuclear pore proteins
49 Nup98 and Nup160 [14-17]. In addition, a screen for sex-peptide insensitive mutants identified
50 Nup54 functionality to be important for neuronal wiring of circuits involved in regulating post-
51 mating behaviours [10]. In particular, a deletion in the promoter of the *Nup54* gene has been
52 associated with altered nucleo-cytoplasmic shuttling in eight *pickpocket* (*ppk)* expressing
53 neurons in the central brain leading to wiring defects and a compromised female post-mating
54 response directed by male-derived sex-peptide transferred during mating [10]. Moreover, this
55 *Nup54* promoter deletion allele maps to a hot spot for rapid evolution in the *Nup54* promotor
56 suggesting sexual conflict driving female escape from male manipulation by sex-peptide under
57 unfavourable conditions [10]. Furthermore, channel Nups have also been attributed a role in
58 transposon silencing in the germline by facilitating processing of short piRNAs from long pre-
59 curser RNAs of the *flamenco* locus, which functions as a 'master off switch' for transposons
60 [18], indicating that the disruption to NPC regulation likely has pleiotropic effects.

61 The nuclear pore protein complex (NPC) provides a physical barrier between nucleus
62 and cytoplasm requiring active transport for cargos above about 40 kDa [19]. The NPC consists
63 of 30 evolutionary highly conserved proteins in *Drosophila* (Fig. 1), which are grouped into
64 subcomplexes termed outer ring (OR), inner ring (IR), cytoplasmic filaments (CF), nuclear
65 basket (NB) and pore membrane proteins (POMS). The transport channel is made up of three
66 phenylalanine-glycine-rich (FG) repeat domains of Nup54, Nup58 and Nup62. The NPC can
67 contribute to differential expression of other genes as e.g. actively transcribed genes can be in
68 the proximity of the NPC [20-23].

69 Given the rapid evolution of *Nup54* gene expression control and the links of Nups to
70 speciation, we systematically examined, whether the NPC is under increased evolutionary
71 constraint as the molecular mechanisms of how the NPC is involved in speciation remains

72 unexplained. Our systematic analysis of Nup coding and promoter regions using *Drosophila*
73 phylogenomics reveals promoters of Nup core genes as hot spots for insertion/deletion (indel)
74 driven rapid evolution and the nuclear pore complex as a nexus for species-specific changes
75 *via* nucleo-cytoplasmic transport regulated gene expression.
76
77 **Results**
78 **Rapid evolution of core nuclear pore protein genes is not generally driven through**
79 **adaptation in coding sequences**
80 To assess whether the protein coding region (Open Reading Frame, ORF) of Nups were
81 under selection, we performed McDonald-Kreitman tests (MKTs) to compare the number of
82 polymorphisms in the *D. melanogaster* ancestral Congo population, and between *D.*
83 *melanogaster* and its closest relative *D. simulans*. This analysis showed that most Nup ORFs
84 do not deviate from neutral evolution (Fig. 1B), but that 30% (3 out of 10) of the Nups
85 constituting the IR showed significant evidence of selection.
86
87 **Promoter regions of channel Nups 58 and 62 have diverged in closely related species**
88 Since we previously identified the promoter of the channel Nup54 as a hot spot for rapid
89 evolution [10], we examined the promoters of the other two channel Nups 58 and 62 among
90 closely related species of *Drosophila* (*D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba*
91 and *D. erecta*) (Fig. 2). This analysis revealed that also the promoters of the other two channel
92 Nups are hot spots for rapid evolution (Fig. 2). Moreover, these changes contain a number of
93 indels (Figure 2C and D) that can fundamentally impact on transcription factor binding
94 commonly found in the proximal region before the TATA box [24-28].
95

96 **Inner and outer ring nuclear pore protein gene subcomplexes undergo rapid evolution**
97 **through indel accumulation in promoters**

98 Next, we examined the promoter region of all remaining Nups in the closely related
99 *Drosophila* species to see whether accumulation of mutations is a general feature of the
100 promoters of this class of genes (Fig. 3, Supplementary Fig. S1). As a control group we used
101 the m^6A mRNA methylation machinery (consisting of Mettl3, Mettl14, fl(2)d, virilizer, flacc,
102 nito and Hakai and the two readers YTHDC1 and YTHDF), because high evolutionary
103 conservation is required to maintain complex stoichiometry to guarantee functionality, making
104 this group optimal for use as a control group to monitor promoter evolution [29-32]. In contrast to
105 the genes from the m^6A mRNA methylation machinery, promoters of ribosomal genes are
106 different in their transcription start site (TSS) properties [33,34].

107 As has been observed for *Drosophila* and human promoters [28,33,35], the region
108 immediately before the TATA box constituting transcription factor binding sites (-30 to -380)
109 shows a slightly increased rate of sequence changes in m6A mRNA methylation writer
110 complex and readers consistent with the general trend (Fig. 3B, Supplementary Fig. 1). We
111 then analysed the promoters of the different Nup sub-complexes IR, OR, CF and NF. This
112 analysis revealed that the core nuclear complex consisting of the IR and OR genes showed a
113 significantly enhanced rate of indel driven evolution compared to the control group or the CF
114 and NB group of genes (Fig. 3 and Supplementary Fig. S1C-F). We next analysed all sequence
115 differences and performed a more detailed analysis split into deletions, insertions and base
116 changes, relative to *D. melanogaster*. Analysis of all sequence differences reflected our indel
117 analysis, with a significant increase in the number of differences in IR and OR Nups
118 (Supplementary Fig. S1A and B). Our more detailed analysis revealed a significant increase in
119 the number of deletions in IR, OR and to a lesser extend in the cytoplasmic filaments (CF)
120 compared to the control (Supplementary Fig. S1C and D), while no significant changes were

121 detected for insertions (Supplementary Fig. S1E and F). In contrast base changes are increased
122 for IR and reduced for OR and CF (Supplementary Fig. S1G and H).
123
124 **Discussion**
125 Here, through comparison of polymorphism rates between closely related Drosophila
126 species we show that rapid evolution of Nups is not generally driven through changes in ORFs.
127 Focussing on upstream and downstream Nup gene regions, we reveal a bias for inclusion of
128 nucleotide changes in the core members of the NPC. Consistent with channel *Nup54* [10], further
129 analysis of upstream regions of the other channel Nups 58 and 62 reveals their promotor regions
130 to be hotspots for rapid indel accumulation. Through examination of promoter regions of Nup
131 subcomplex members, we identify the core IR and OR Nups as hot spots for rapid accumulation
132 of indels. Here, the core NPC members likely act as drivers of species-specific variation
133 through dominant indel driven changes in promoters, subsequently modifying regulation of
134 nucleo-cytoplasmic transport regulated gene expression.
135 Although the megadalton nuclear pore protein complex (NPC) has mainly been
136 associated with providing a physical barrier between nucleus and cytoplasm, recent discovery
137 of pleiotropic functions of channel Nup54 provide new insights how the NPC could drive
138 speciation. Nup54 is required for neuronal wiring underlying the female post-mating response
139 induced by male-derived sex-peptide. Since its promoter is a hot spot for rapid evolution, a
140 critical role for regulatory elements at the onset of speciation is indicated as a result of sexual
141 conflict [5,9,10,13,36].
142 Given the role of channel Nups in piRNA meditated transposon silencing of the
143 *flamenco* locus in the germline and their requirement for sexual differentiation [10,18], changes at
144 one trait could cascade into the second leading to rapid phenotypic evolution and speciation.

145 Intriguingly, mutations in *cis*-regulatory sequences of genes are often dominant whereas the
146 majority of coding mutations are recessive [37-39] leading to different efficacy of selection [40-42].
147 Maternally inherited piRNAs are essential to transposon silencing and an imbalance
148 can lead to a phenomenon called hybrid dysgenesis, imposing reduced fecundity in the female
149 offspring when the male genome contributes novel transposable elements to be silenced [43]. Key
150 to silencing are epigenetically inherited piRNAs from the female to prime the "ping-pong"
151 cycle for amplification to increase the silencing capacity by heterochromatinization for
152 preventing transposon mobilisation [43-46]. Compromised transposon silencing leading to
153 deleterious reductions of fecundity could be escaped by altered NPC function. Such changes
154 are likely to trigger pleiotropic effects including altered neuronal wiring and change of
155 behaviour, leading to accelerated evolution and speciation. In essence, if escape from lower
156 fecundity as a result of hybrid dysgenesis [43-46], leading to reduced transposon silencing is
157 coupled to changes in the morphology or behaviour, new species could be established very
158 quickly. In particular, if changes in Nup functionality result in rewiring of neuronal circuits as
159 indicated for Nup54, behavioural preferences could lead to rapid isolation. In fact, mate
160 preference has changed in *D. simulans* through rewiring of sensory neuron projections to
161 fruitless P1 neurons that control courtship [47].
162
163 **Conclusion**
164 Changes in gene expression have profound effects during species divergence and
165 phenotypic adaptation, and such changes can lead to hybrids' gene mis-expression and
166 dysfunction [48-52]. The molecular mechanism how new species arise through differential
167 regulation of gene expression remains uncertain. The newly discovered role of channel Nups
168 in piRNA processing in the germline to maintain transposon silencing provokes the claim that
169 any compensation of the negative impact on fecundity from hybrid dysgenesis would be

170 favoured. However, due to the pleiotropic effect of Nups on neuronal wiring and behaviour,
171 the escape from hybrid incompatibility could produce changes leading to behavioural isolation
172 [10,18]. Since mutations in promoters can exert a dominant effect [37-39], such changes could be
173 rapidly fixed in contrast to recessive changes in coding regions, which would remain hidden in
174 heterozygosity. Our systematic analysis of the evolution of all Nups coding and promoter
175 regions suggests a less studied mode of evolution through changes of sequences upstream of
176 the transcription start site (TSS), particularly in the promoters of core Nups, and supports the
177 possibility that compensation of deleterious changes in the germline can lead to altered
178 neuronal wiring and rapid fixation of adaptive traits.
179

180 **Materials and Methods**

181 **Open reading frame analysis**

182 To determine whether the ORFs of nucleoporins are under selection, the PopFly online
183 database (imkt.uab.cat) developed from the Drosophila Genome Nexus project assembling
184 sequence data of around 1100 *D. melanogaster* genomes were used to perform MKTs to
185 analyse polymorphism data from the ancestral Congo population because it is a sub-Saharan
186 population with higher ancestral stability than other populations [53-55]. Synonymous and non-
187 synonymous polymorphisms within *D. melanogaster* or between *D. melanogaster* and *D.*
188 *simulans* were obtained and significance determined by Fisher's Exact Test, with significance
189 defined as $p \leq 0.05$ with Bonferroni correction.
190

191 **Sequence retrieval and alignment**

192 Nucleoporin gene and promoter sequences were retrieved from UCSC genome browser
193 (genome.ucsc.edu) by entering the species and gene name, and then define the gene region

194 under tools/table browser. Alignments were done with clustalW within the MEGA11 program
195 [56].
196
197 **Accumulation of substitutions along extended gene regions**
198 To test for nonrandom accumulation of substitutions along the Nup extended gene
199 regions we determined significant deviations from a uniform distribution of substitutions using
200 an empirical cumulative distribution function. The function (G) detects monotonic increases in
201 substitutions (n) measured as the difference between the relative occurrence of a nucleotide
202 change and its relative position in the alignment [57]. Whether differences between the values of
203 the G function (ΔG) between substitutional events deviates from a random accumulation of
204 changes is tested using Monte Carlo simulations to produce 100,000 samples of n events by
205 sampling sites without replacement along the alignment [57].
206
207 **Comparison of substitution rates**
208 To analyse the promoters of Nups and the control group of the m6A writer complex
209 and readers a region of 1000 bp upstream and downstream of the transcriptional start site (TSS)
210 was used with TSS as an anchor point. Regions of differential alignment between the
211 *Drosophila* species were averaged across the NPC subcomplexes, where alignments were
212 translated into events [58]. Per [58], 0 signified no sequence difference between all analysed species
213 and 1 signified comparative sequence divergence for ≥1 species. A sliding window of five
214 nucleotides upstream and downstream of each position was summed and processed to define
215 sliding event (*Se*) scores which were used to generate heatmaps. To calculate sequence change
216 accumulation (percentage of events greater than the average control promoter sliding window
217 score (*d*)), a 350-nucleotide region upstream of the approximated TATA box region was taken.

218 The total number of *Se* scores greater than the control *Se* (Se^C) was divided by the total number

219 of events in the region (*N*).

220 $$d = \frac{\text{Total number of } Se \text{ events where } Se > Se^C}{N} \text{x}100$$

221 Significance was defined by non-parametric chi-squared tests vs the control score. P values

222 ≤0.05 were deemed statistically significant.

223

224 **Data availability**

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

226

227 **Acknowledgements**

228 This work was supported by the Medical Research Council Integrated Midlands Partnership

229 for Biomedical Training Program by D.W.J.M., the Biotechnology and Biological Science

230 Research Council to M.S. and the Natural Sciences and Engineering Research Council of

231 Canada to A.C. We thank R. Arnold for help with bioinformatics.

232

233 **Authors contributions**

234 M.S. conceived and directed the project. All authors analysed data. M.S., D.W.J.M. and A.C.

235 wrote the manuscript. All authors read and approved the final manuscript.

236

237 **Competing interests**

238 The authors declare no competing interests.

239

240 **Additional information**

241 Supplementary information includes Fig. S1 and Supplementary Dataset 1.

242

243 **References**

244 1 Avila, F. W., Sirot, L. K., LaFlamme, B. A., Rubinstein, C. D. & Wolfner, M. F.
245 Insect seminal fluid proteins: identification and function. *Annu Rev Entomol* **56**, 21-
246 40, doi:10.1146/annurev-ento-120709-144823 (2011).

247 2 Gillott, C. Male accessory glandd secretions: Modulators of Female Reproductive
248 Physiology and Behavior. *Annu. Rev. Entomol.* **48**, 163-184 (2003).

249 3 Kubli, E. Sex-peptides: seminal peptides of the Drosophila male. *Cell Mol Life Sci* **60**,
250 1689-1704 (2003).

251 4 Soller, M., Bownes, M. & Kubli, E. Control of oocyte maturation in sexually mature
252 Drosophila females. *Dev Biol* **208**, 337-351 (1999).

253 5 Chapman, T. Sexual conflict: mechanisms and emerging themes in resistance biology.
254 *The American Naturalist* **192**, 217-229 (2018).

255 6 Coyne, J. A. & Orr, H. A. *Speciation*. (Sinauer Associates, 2004).

256 7 Grayson, P. Izumo1 and Juno: the evolutionary origins and coevolution of essential
257 sperm-egg binding partners. *Royal Society Open Science* **2**, 150296,
258 doi:doi:10.1098/rsos.150296 (2015).

259 8 Hollis, B. *et al.* Sexual conflict drives male manipulation of female postmating
260 responses in Drosophila melanogaster. *Proc Natl Acad Sci U S A* **116**, 8437-8444,
261 doi:10.1073/pnas.1821386116 (2019).

262 9 Hopkins, B. R. & Perry, J. C. The evolution of sex peptide: sexual conflict,
263 cooperation, and coevolution. *Biological Reviews*, doi:10.1111/brv.12849 (2022).

264 10 Nallasivan, M. P., Haussmann, I. U., Civetta, A. & Soller, M. Channel nuclear pore
265 protein 54 directs sexual differentiation and neuronal wiring of female reproductive
266 behaviors in Drosophila. *BMC biology* **19**, 1-14 (2021).

267 11 Patlar, B., Jayaswal, V., Ranz, J. M. & Civetta, A. Nonadaptive molecular evolution
268 of seminal fluid proteins in Drosophila. *Evolution; International Journal of Organic*
269 *Evolution* **75**, 2102 - 2113 (2021).
270 12 Rowe, M. *et al.* Molecular Diversification of the Seminal Fluid Proteome in a
271 Recently Diverged Passerine Species Pair. *Mol Biol Evol* **37**, 488-506,
272 doi:10.1093/molbev/msz235 (2020).
273 13 Sirot, L. K., Wong, A., Chapman, T. & Wolfner, M. F. Sexual conflict and seminal
274 fluid proteins: a dynamic landscape of sexual interactions. *Cold Spring Harb Perspect*
275 *Biol* **7**, a017533, doi:10.1101/cshperspect.a017533 (2014).
276 14 Presgraves, D. C. The molecular evolutionary basis of species formation. *Nature*
277 *Reviews Genetics* **11**, 175-180 (2010).
278 15 Presgraves, D. C., Balagopalan, L., Abmayr, S. M. & Orr, H. A. Adaptive evolution
279 drives divergence of a hybrid inviability gene between two species of Drosophila.
280 *Nature* **423**, 715-719 (2003).
281 16 Tang, S. & Presgraves, D. C. Evolution of the Drosophila nuclear pore complex
282 results in multiple hybrid incompatibilities. *Science* **323**, 779-782 (2009).
283 17 Tang, S. & Presgraves, D. C. Lineage-specific evolution of the complex Nup160
284 hybrid incompatibility between Drosophila melanogaster and its sister species.
285 *Genetics* **200**, 1245-1254 (2015).
286 18 Munafo, M. *et al.* Channel nuclear pore complex subunits are required for transposon
287 silencing in Drosophila. *Elife* **10**, e66321, doi:10.7554/eLife.66321 (2021).
288 19 Lin, D. H. & Hoelz, A. The Structure of the Nuclear Pore Complex (An Update).
289 *Annu Rev Biochem* **88**, 18.11-18.59, doi:10.1146/annurev-biochem-062917-011901
290 (2019).

291 20 Cabal, G. G. *et al.* SAGA interacting factors confine sub-diffusion of transcribed
292 genes to the nuclear envelope. *Nature* **441**, 770-773 (2006).
293 21 Pascual-Garcia, P. & Capelson, M. Nuclear pores as versatile platforms for gene
294 regulation. *Current opinion in genetics & development* **25**, 110-117 (2014).
295 22 Taddei, A. *et al.* Nuclear pore association confers optimal expression levels for an
296 inducible yeast gene. *Nature* **441**, 774-778 (2006).
297 23 Breuer, M. & Ohkura, H. A negative loop within the nuclear pore complex controls
298 global chromatin organization. *Genes & Development* **29**, 1789-1794,
299 doi:10.1101/gad.264341.115 (2015).
300 24 Hoogendoorn, B. *et al.* Functional analysis of human promoter polymorphisms.
301 *Human Molecular Genetics* **12**, 2249-2254, doi:10.1093/hmg/ddg246 (2003).
302 25 Smale, S. T. & Kadonaga, J. T. The RNA polymerase II core promoter. *Annu Rev*
303 *Biochem* **72**, 449-479, doi:10.1146/annurev.biochem.72.121801.161520 (2003).
304 26 Vo Ngoc, L., Kassavetis, G. A. & Kadonaga, J. T. The RNA Polymerase II Core
305 Promoter in Drosophila. *Genetics* **212**, 13-24, doi:10.1534/genetics.119.302021
306 (2019).
307 27 Hill, M. S., Vande Zande, P. & Wittkopp, P. J. Molecular and evolutionary processes
308 generating variation in gene expression. *Nature Reviews Genetics* **22**, 203-215,
309 doi:10.1038/s41576-020-00304-w (2021).
310 28 Hoskins, R. A. *et al.* Genome-wide analysis of promoter architecture in Drosophila
311 melanogaster. *Genome Research* **21**, 182-192, doi:10.1101/gr.112466.110 (2011).
312 29 Balacco, D. L. & Soller, M. The m(6)A Writer: Rise of a Machine for Growing Tasks.
313 *Biochemistry* **58**, 363-378, doi:10.1021/acs.biochem.8b01166 (2019).

314 30 Bawankar, P. *et al.* Hakai is required for stabilization of core components of the
315 m(6)A mRNA methylation machinery. *Nat Commun* **12**, 3778, doi:10.1038/s41467-
316 021-23892-5 (2021).
317 31 Dezi, V., Ivanov, C., Haussmann, I. U. & Soller, M. Nucleotide modifications in
318 messenger RNA and their role in development and disease. *Biochem Soc Trans* **44**,
319 1385-1393, doi:10.1042/bst20160110 (2016).
320 32 Knuckles, P. *et al.* Zc3h13/Flacc is required for adenosine methylation by bridging
321 the mRNA-binding factor Rbm15/Spenito to the m(6)A machinery component
322 Wtap/Fl(2)d. *Genes Dev* **32**, 415-429, doi:10.1101/gad.309146.117 (2018).
323 33 Main, B. J., Smith, A. D., Jang, H. & Nuzhdin, S. V. Transcription Start Site
324 Evolution in Drosophila. *Molecular Biology and Evolution* **30**, 1966-1974,
325 doi:10.1093/molbev/mst085 (2013).
326 34 Anreiter, I., Tian, Y. W. & Soller, M. The cap epitranscriptome: Early directions to a
327 complex life as mRNA. *BioEssays*, 2200198, doi:10.1002/bies.202200198 (2022).
328 35 Taylor, M. S. *et al.* Heterotachy in mammalian promoter evolution. *PLoS genetics* **2**,
329 e30 (2006).
330 36 Hollis, B. *et al.* Sexual conflict drives male manipulation of female postmating
331 responses in Drosophila melanogaster. *Proceedings of the National Academy of*
332 *Sciences* **116**, 8437-8444 (2019).
333 37 Wittkopp, P. J., Haerum, B. K. & Clark, A. G. Evolutionary changes in cis and trans
334 gene regulation. *Nature* **430**, 85-88 (2004).
335 38 Pastinen, T. *et al.* A survey of genetic and epigenetic variation affecting human gene
336 expression. *Physiological genomics* (2004).
337 39 Ronald, J., Brem, R. B., Whittle, J. & Kruglyak, L. Local regulatory variation in
338 Saccharomyces cerevisiae. *PLoS genetics* **1**, e25 (2005).

339 40 Wilkins, A. S. *The evolution of developmental pathways*. (Sinauer Associates, 2002).
340 41 Stern, D. L. Perspective: evolutionary developmental biology and the problem of
341 variation. *Evolution* **54**, 1079-1091 (2000).
342 42 Wray, G. A. *et al.* The evolution of transcriptional regulation in eukaryotes. *Mol Biol*
343 *Evol* **20**, 1377-1419, doi:10.1093/molbev/msg140 (2003).
344 43 Czech, B. *et al.* piRNA-Guided Genome Defense: From Biogenesis to Silencing.
345 *Annu Rev Genet* **52**, 131-157, doi:10.1146/annurev-genet-120417-031441 (2018).
346 44 Blumenstiel, J. P. & Hartl, D. L. Evidence for maternally transmitted small interfering
347 RNA in the repression of transposition in Drosophila virilis. *Proc Natl Acad Sci U S A*
348 **102**, 15965-15970, doi:10.1073/pnas.0508192102 (2005).
349 45 Malone, C. D. & Hannon, G. J. Molecular evolution of piRNA and transposon control
350 pathways in Drosophila. *Cold Spring Harb Symp Quant Biol* **74**, 225-234,
351 doi:10.1101/sqb.2009.74.052 (2009).
352 46 Brennecke, J. *et al.* An epigenetic role for maternally inherited piRNAs in transposon
353 silencing. *Science* **322**, 1387-1392, doi:10.1126/science.1165171 (2008).
354 47 Seeholzer, L. F., Seppo, M., Stern, D. L. & Ruta, V. Evolution of a central neural
355 circuit underlies Drosophila mate preferences. *Nature* **559**, 564-569,
356 doi:10.1038/s41586-018-0322-9 (2018).
357 48 Cutter, A. D. & Bundus, J. D. Speciation and the developmental alarm clock. *Elife* **9**,
358 doi:10.7554/eLife.56276 (2020).
359 49 Hagen, J. F. D. *et al.* tartan underlies the evolution of Drosophila male genital
360 morphology. *Proceedings of the National Academy of Sciences of the United States of*
361 *America* **116**, 19025-19030, doi:10.1073/pnas.1909829116 (2019).
362 50 Mack, K. L. & Nachman, M. W. Gene Regulation and Speciation. *Trends Genet* **33**,
363 68-80, doi:10.1016/j.tig.2016.11.003 (2017).

364 51 Civetta, A. Misregulation of Gene Expression and Sterility in Interspecies Hybrids:
365 Causal Links and Alternative Hypotheses. *J Mol Evol* **82**, 176-182,
366 doi:10.1007/s00239-016-9734-z (2016).
367 52 King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees.
368 *Science* **188**, 107-116, doi:10.1126/science.1090005 (1975).
369 53 Lack, J. B. *et al.* The Drosophila genome nexus: a population genomic resource of
370 623 Drosophila melanogaster genomes, including 197 from a single ancestral range
371 population. *Genetics* **199**, 1229-1241, doi:10.1534/genetics.115.174664 (2015).
372 54 Lack, J. B., Lange, J. D., Tang, A. D., Corbett-Detig, R. B. & Pool, J. E. A Thousand
373 Fly Genomes: An Expanded Drosophila Genome Nexus. *Mol Biol Evol* **33**, 3308-
374 3313, doi:10.1093/molbev/msw195 (2016).
375 55 Mackay, T. F. *et al.* The Drosophila melanogaster Genetic Reference Panel. *Nature*
376 **482**, 173-178, doi:10.1038/nature10811 (2012).
377 56 Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular
378 Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol* **35**, 1547-
379 1549, doi:10.1093/molbev/msy096 (2018).
380 57 Civetta, A., Ostapchuk, D. C. & Nwali, B. Genome Hotspots for Nucleotide
381 Substitutions and the Evolution of Influenza A (H1N1) Human Strains. *Genome Biol*
382 *Evol* **8**, 986-993, doi:10.1093/gbe/evw061 (2016).
383 58 Tang, H. & Lewontin, R. C. Locating Regions of Differential Variability in DNA and
384 Protein Sequences. *Genetics* **153**, 485-495, doi:10.1093/genetics/153.1.485 (1999).
385

386 **Figure legends**

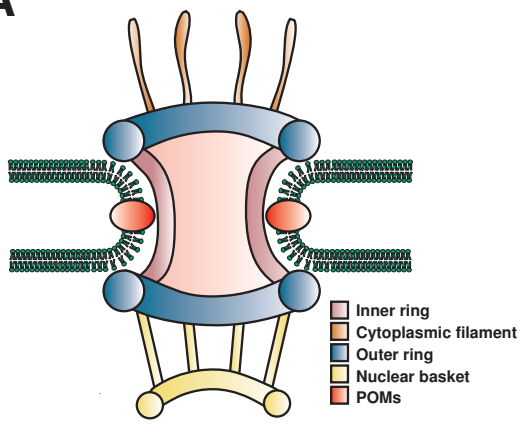
387 **Fig. 1: Core nuclear pore protein genes do not generally evolve through adaptation in**
388 **coding sequences (ORFs).**

389 **A**) Graphic depiction of the nuclear pore complex with differentially coloured subcomplexes
390 inner ring (IR, pink), outer ring (OR, blue), cytoplasmic filament (CF, orange), nuclear basket
391 (yellow) and pore membrane proteins (POMS, red) embedded in the nuclear envelope lipid-
392 bilayer (green).
393 **B)** Significance of synonymous to non-synonymous changes in open reading frames (ORFs)
394 indicated by an asterisk and p-values.
395
396 **Fig. 2: Promoter regions of channel Nups 58 and 62 have diverged in closely related**
397 **species.**
398 **A, B)** Sequence alignment of the Nup58 and Nup62 promoter regions from closely related
399 species. Nucleic acids changes from *D. melanogaster* are indicated in black. Transcribed parts
400 of the *Vha13* and *CG7997* 3′UTR and the *Nup58* and *Nup62* 5′UTR are indicated by a line.
401 **C, D)** Plot of cumulative differences along the sequence (G) between the relative occurrences
402 of indels and their position from the alignment of the gene region around *Nup58* and *Nup62*
403 between *D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba* and *D. erecta*. Positions in the
404 alignment with significant stretches of substitutions are identified by black line(s).
405
406 **Fig. 3: Core nuclear pore protein genes rapidly evolve through indel accumulation in**
407 **promoters.**
408 **A)** Heatmaps indicating divergence for indels in yellow among closely related *D.*
409 *melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba* and *D. erecta* for the control group of genes
410 (m6A writer complex and readers) and the inner ring (IR), outer ring (OR), cytoplasmic
411 filament (CF) and nuclear basket (NB) below the gene model with the transcription start site
412 (TSS) indicated by an arrow. The red line indicates the promoter region used for quantification
413 of the substitution rate in D.

414 **B)** Quantification of the change rate of indels in the TATA box distal region (from -30 to -
415 380). Statistically significant differences from non-parametric chi-squared tests are indicated
416 by asterisks (* $p \leq 0.05$, ** $p \leq 0.001$, *** $p \leq 0.0001$).
417
418 **Supplementary Fig. S1: Inner and outer ring nuclear pore protein gene subcomplexes**
419 **undergo rapid evolution through indel accumulation in promoters.**
420 **A)** Gene model depicting the intergenic region followed by the transcription start site (TSS)
421 and the genespan with the 350 bp region analysed for promotor activity upstream of the TATA
422 box location.
423 **A-H)** Heatmaps indicating divergence in yellow among closely related Drosophila species for
424 the control group of genes (m6A writer complex and readers) and the inner ring (IR), outer ring
425 (OR), cytoplasmic filament (CF) and nuclear basket (NB). Calculated sequence difference
426 index, deletion index, insertion index and base change index events are reported as bar charts
427 for the NPC subgroups compared to the control group. Index events were defined by values
428 above the control intergenic region average. Statistically significant differences between the
429 NPC subgroups and the control group were calculated from chi-squared tests and indicated by
430 asterisks (* $p \leq 0.05$, ** $p \leq 0.001$, *** $p \leq 0.0001$).
431

Fig. 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

Fig. 2

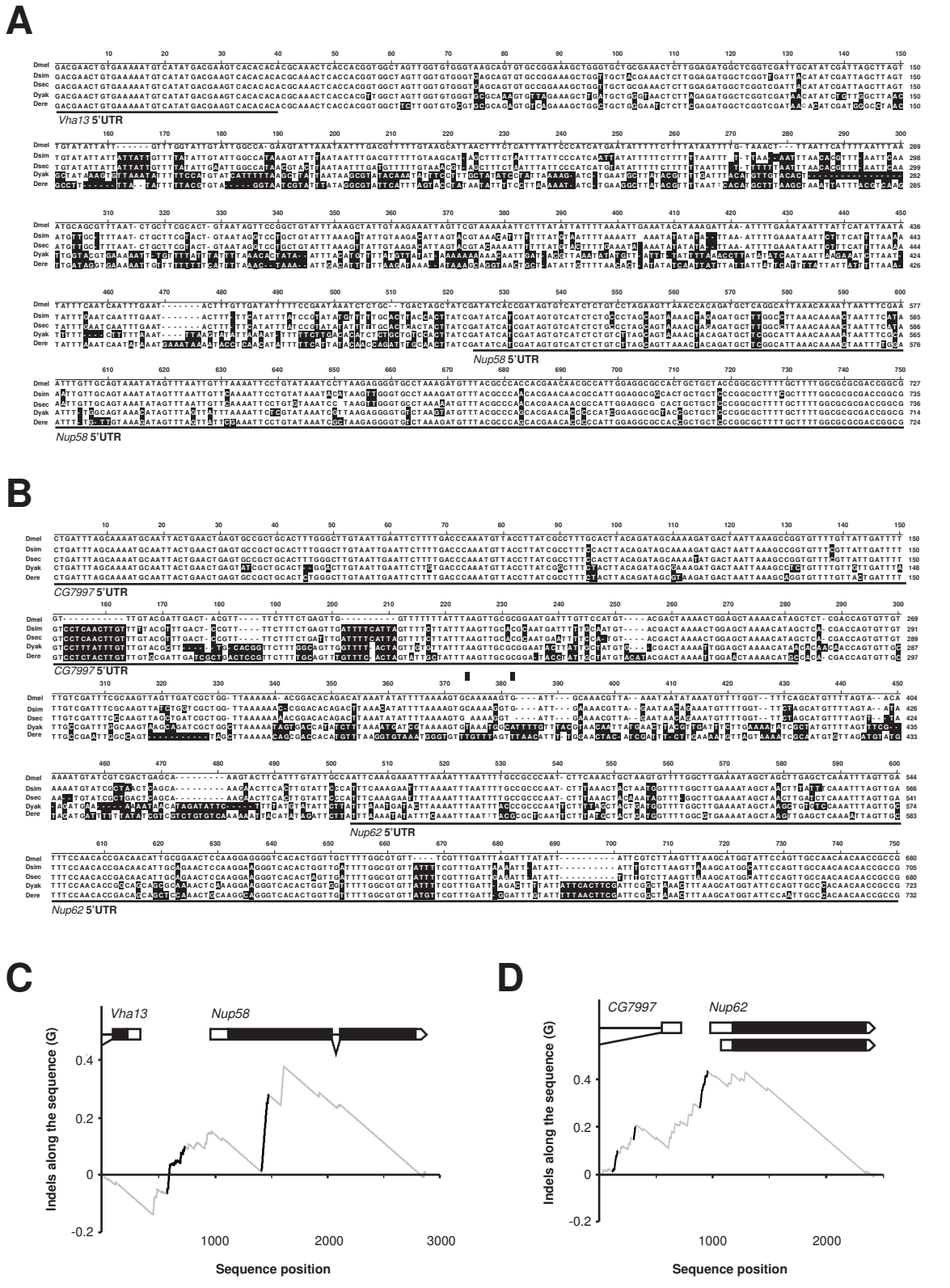
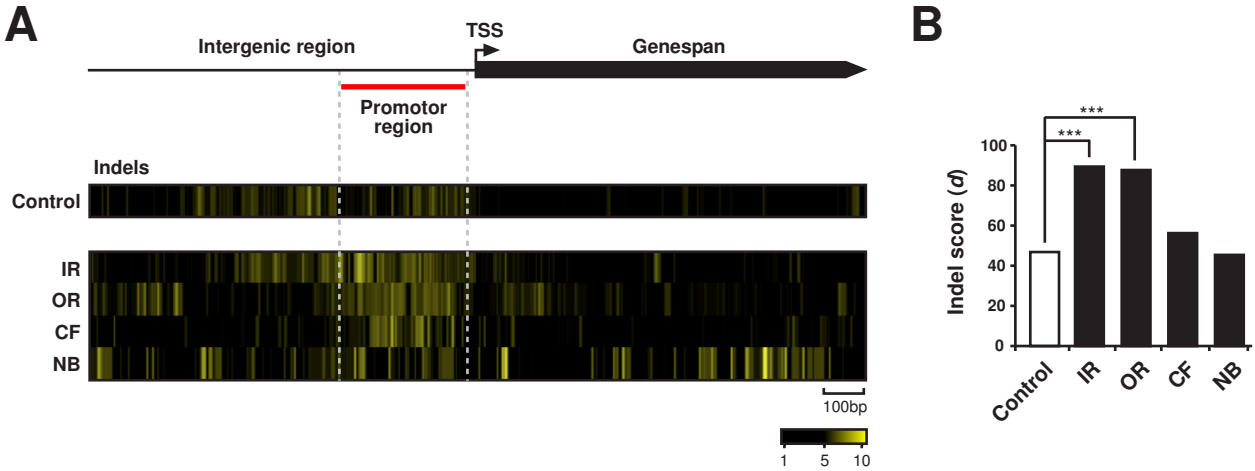
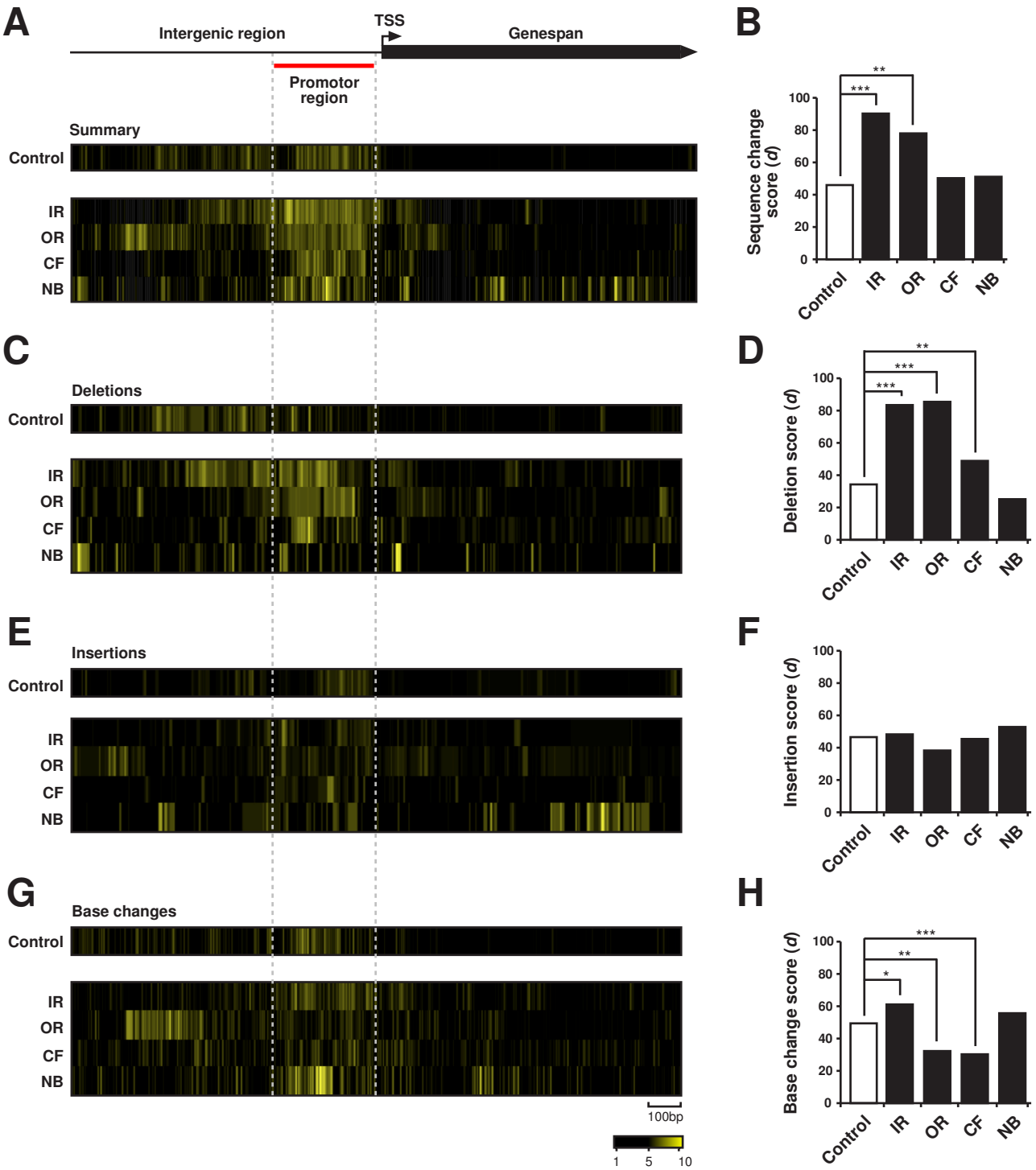

A
Dmel
Dsim
Dsec
Dyak
Dere
Vha13 5'UTR
Nup58 5'UTR
B
CG7997 5'UTR
Nup62 5'UTR
C
Vha13
Nup58
Indels along the sequence (G)
Sequence position
D
CG7997
Nup62
Indels along the sequence (G)
Sequence position

Fig. 3


A
Intergenic region
TSS
Genespan
Promotor region
Indels
Control
IR
OR
CF
NB
100bp
1
5
10
B
***
***
Indel score (d)
100
80
60
40
20
0
Control
IR
OR
CF
NB

Supplementary Fig. S1