Indel driven rapid evolution of core nuclear pore protein gene promoters

ADAM M. READ[1*], DAVID W. J. MCQUARRIE[1,2*], 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, 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 polymorphisms 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*, the other 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. Since changes in promoters have dominant effects, these results indicate a novel 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.

Significance

Nuclear pore proteins (Nups) have been implicated as drivers of speciation in *Drosophila*, though the mechanism driving this effect is unclear. Here, we reveal indel accumulation in promoter regions as a potential mode of rapid evolution shared between the channel Nups, and Nups comprising the nuclear pore complex inner and outer rings. Since changes in promoters have dominant effects, these results highlight a less explored mode of

47 rapid evolution through indel accumulation in Nup inner and outer ring promoters which could
48 provide a mechanism for Nup-driven speciation.
49

50 **Introduction**

51 Male derived substances transferred during mating change female physiology and
52 behaviour to guarantee reproductive success in many insects (Avila, et al. 2011; Gillott 2003;
53 Kubli 2003; Soller, et al. 1999). Rapid divergence in male versus female interactive molecules
54 can hamper reproductive success, potentially leading to the establishment of new species for
55 example by compromising fertility or viability among hybrids (Chapman 2018; Coyne and Orr
56 2004; Grayson 2015; Hollis, et al. 2019a; Hopkins and Perry 2022; Nallasivan, et al. 2021;
57 Patlar, et al. 2021; Rowe, et al. 2020; Sirot, et al. 2014). Such hybrid incompatibility among
58 closely related *Drosophila* species has been used to identify responsible genes. Among the few
59 genes identified were unexpectedly nuclear pore proteins Nup98 and Nup160 (Presgraves
60 2010; Presgraves, et al. 2003; Tang and Presgraves 2009, 2015). In addition, a screen for sex-
61 peptide insensitive mutants identified Nup54 functionality to be important for neuronal wiring
62 of circuits involved in regulating post-mating behaviours (Nallasivan, et al. 2021).
63 Additionally, channel Nups have also been attributed a role in transposon silencing in the
64 germline by facilitating processing of short piRNAs from long pre-curser RNAs of the
65 *flamenco* locus (Munafo, et al. 2021), indicating that the disruption to NPC regulation likely
66 has pleiotropic effects.

67 The nuclear pore protein complex (NPC) provides a physical barrier between nucleus
68 and cytoplasm requiring active transport for cargos above about 40 kDa (Lin and Hoelz 2019).
69 The NPC consists of 30 evolutionary highly conserved proteins in *Drosophila* (fig. 1), which
70 are grouped into subcomplexes termed outer ring (OR), inner ring (IR), cytoplasmic filaments
71 (CF), nuclear basket (NB) and pore membrane proteins (POMS). The transport channel is made

72 up of three phenylalanine-glycine-rich (FG) repeat domains of Nup54, Nup58 and Nup62. The
73 NPC can contribute to differential expression of other genes as e.g. actively transcribed genes
74 can be in the proximity of the NPC (Cabal, et al. 2006; Pascual-Garcia and Capelson 2014;
75 Taddei, et al. 2006).

76 However, whether the NPC is under evolutionary constraint has not been systematically
77 examined and the molecular mechanism of how the NPC is involved in speciation remains
78 unexplained. Our systematic analysis of Nup coding and promoter regions using *Drosophila*
79 phylogenomics reveals promoters of Nup core genes as hot spots for insertion/deletion (indel)
80 driven rapid evolution and the nuclear pore complex as a nexus for species-specific changes
81 *via* nucleo-cytoplasmic transport regulated gene expression.

82

83 **Results**

84 **Rapid Evolution of Core Nuclear Pore Protein Genes is Not Generally Driven Through**
85 **Adaptation in Coding Sequences**

86 To assess whether the protein coding region (Open Reading Frame, ORF) of Nups were
87 under selection, we compared the number of polymorphisms between individual *D.*
88 *melanogaster* lines and between *D. melanogaster* and its closest relative *D. simulans*. This
89 analysis showed that most Nup ORFs do not deviate from neutral evolution (fig. 1B), but that
90 60% (6 out of 10) of the Nups constituting the IR showed significant evidence of selection. In
91 addition to the previously identified OR Nups 96 and 160, and also Nup75 showed evidence
92 of evolving under selection, but this represents only 30 % (3 out of 10) of the OR.

93

94 **Promoter Regions of Channel Nups 58 and 62 Have Diverged in Closely Related Species**

95 Since we previously identified the promoter of the channel Nup54 as a hot spot for rapid
96 evolution (Nallasivan, et al. 2021), we examined the promoters of the other two channel Nups

97 58 and 62 among closely related species of Drosophila (*D. melanogaster*, *D. simulans*, *D.*
98 *sechellia*, *D. yakuba* and *D. erecta*) (fig. 2). This analysis revealed that also the promoters of
99 the other two channel Nups are hot spots for rapid evolution (fig. 2). Moreover, these changes
100 contain a number of indels (Figure 2C and D) that can fundamentally impact on transcription
101 factor binding commonly found in the proximal region before the TATA box (Hill, et al. 2021;
102 Hoogendoorn, et al. 2003; Smale and Kadonaga 2003; Vo Ngoc, et al. 2019).
103

104 **Inner and Outer Ring Nups Have Significantly Increased Sequence Changes in Gene**
105 **Upstream Regions**

106 We next assessed the accumulation of nucleotide changes and indels among closely
107 related species of Drosophila (*D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba* and *D.*
108 *erecta*) by comparing the proportion of indel events and nucleotide distance. This analysis
109 showed that on average, the sequence upstream of the transcription start site (TSS) of Nups,
110 where putative regulatory elements are located, have evolved faster than the downstream
111 regions corresponding to the gene transcripts. This is particularly the case for Nups of the IR
112 and OR and particularly for indels (Supplementary fig. S1).
113

114 **Inner and Outer Ring Nuclear Pore Protein Gene Subcomplexes Undergo Rapid**
115 **Evolution Through Indel Accumulation in Promoters**

116 Next, we examined the promoter region of all remaining Nups in the closely related
117 Drosophila species to see whether accumulation of mutations is a general feature of the
118 promoters of this class of genes (fig. 3, Supplementary fig. S2). As a control group we used the
119 highly conserved m6A mRNA methylation complex consisting of Mettl3, Mettl14, fl(2)d,
120 virilizer, flacc, nito and hakai and the two readers YTHDC1 and YTHDF (Balacco and Soller
121 2019; Bawankar, et al. 2021; Dezi, et al. 2016; Knuckles, et al. 2018). As has been observed

122 for human promoters (Taylor, et al. 2006), the region immediately before the TATA box
123 constituting transcription factor binding sites (-30 to -380) shows a slightly increased rate of
124 sequence changes in m6A mRNA methylation writer complex and readers (fig. 3B). We then
125 analysed the promoters of the different Nup sub-complexes IR, OR, CF and NF. This analysis
126 revealed that the core nuclear complex consisting of the IR and OR genes showed a
127 significantly enhanced rate of indel driven evolution compared to the control group or the CF
128 and NB group of genes (fig. 3 and Supplementary fig. S2C-F). We next analysed all sequence
129 differences and performed a more detailed analysis split into deletions, insertions and base
130 changes, relative to *D. melanogaster*. Analysis of all sequence differences reflected our indel
131 analysis, with a significant increase in the number of differences in IR and OR Nups
132 (Supplementary fig. S2A and B). Our more detailed analysis revealed a significant increase in
133 the number of deletions in IR, OR and to a lesser extend in the cytoplasmic filaments (CF)
134 compared to the control (Supplementary fig. S2C and D), while no significant changes were
135 detected for insertions (Supplementary fig. S2E and F). In contrast base changes are increased
136 for IR and reduced for OR and CF (Supplementary fig. S2G and H).
137
138 **Discussion**
139 Here, through comparison of polymorphism rates between closely related Drosophila
140 species we show that rapid evolution of Nups is not generally driven through changes in ORFs.
141 Focussing on upstream and downstream Nup gene regions, we reveal a bias for inclusion of
142 nucleotide changes in the core members of the NPC. Consistent with channel *Nup54*, further
143 analysis of upstream regions of the other channel Nups 58 and 62 reveals their promotor regions
144 to be hotspots for rapid indel accumulation. Through examination of promoter regions of Nup
145 subcomplex members, we identify the core IR and OR Nups as hot spots for rapid accumulation
146 of indels. Here, the core NPC members likely act as drivers of species-specific variation

147 through dominant indel driven changes in promoters, subsequently modifying regulation of
148 nucleo-cytoplasmic transport regulated gene expression.

149 Although the megadalton nuclear pore protein complex (NPC) has mainly been
150 associated with providing a physical barrier between nucleus and cytoplasm, recent discovery
151 of pleiotropic functions of channel Nup54 provide new insights how the NPC could drive
152 speciation. Nup54 is required for neuronal wiring underlying the female post-mating response
153 induced by male-derived sex-peptide. Since its promoter is a hot spot for rapid evolution, a
154 critical role for regulatory elements at the onset of speciation is indicated as a result of sexual
155 conflict (Chapman 2018; Hollis, et al. 2019b; Hopkins and Perry 2022; Nallasivan, et al. 2021;
156 Sirot, et al. 2014).

157 Given the role of channel Nups in piRNA meditated transposon silencing of the
158 *flamenco* locus in the germline and their requirement for sexual differentiation (Munafo, et al.
159 2021; Nallasivan, et al. 2021), changes at one trait could cascade into the second leading to
160 rapid phenotypic evolution and speciation. Intriguingly, mutations in *cis*-regulatory sequences
161 of genes are often dominant whereas the majority of coding mutations are recessive (Pastinen,
162 et al. 2004; Ronald, et al. 2005; Wittkopp, et al. 2004) leading to different efficacy of selection
163 (Stern 2000; Wilkins 2002; Wray, et al. 2003).

164 Maternally inherited piRNAs are essential to transposon silencing and an imbalance
165 can lead to a phenomenon called hybrid dysgenesis, imposing reduced fecundity in the female
166 offspring when the male genome contributes novel transposable elements to be silenced
167 (Czech, et al. 2018). Key to silencing are epigenetically inherited piRNAs from the female to
168 prime the "ping-pong" cycle for amplification to increase the silencing capacity by
169 heterochromatinization for preventing transposon mobilisation (Blumenstiel and Hartl 2005;
170 Brennecke, et al. 2008; Czech, et al. 2018; Malone and Hannon 2009). Compromised
171 transposon silencing leading to deleterious reductions of fecundity could be escaped by altered

172 NPC function. Such changes are likely to trigger pleiotropic effects including altered neuronal
173 wiring and change of behaviour, leading to accelerated evolution and speciation. In essence, if
174 escape from lower fecundity as a result of hybrid dysgenesis (Blumenstiel and Hartl 2005;
175 Brennecke, et al. 2008; Czech, et al. 2018; Malone and Hannon 2009), leading to reduced
176 transposon silencing is coupled to changes in the morphology or behaviour, new species could
177 be established very quickly. In particular, if changes in Nup functionality result in rewiring of
178 neuronal circuits as indicated for Nup54, behavioural preferences could lead to rapid isolation.
179 In fact, mate preference has changed in *D. simulans* through rewiring of sensory neuron
180 projections to fruitless P1 neurons that control courtship (Seeholzer, et al. 2018).
181

182 **Conclusion**

183 Changes in gene expression have profound effects during species divergence and
184 phenotypic adaptation, and such changes can lead to hybrids' gene mis-expression and
185 dysfunction (Civetta 2016; Cutter and Bundus 2020; Hagen, et al. 2019; King and Wilson 1975;
186 Mack and Nachman 2017). The molecular mechanism how new species arise through
187 differential regulation of gene expression remains uncertain. The newly discovered role of
188 channel Nups in piRNA processing in the germline to maintain transposon silencing provokes
189 the claim that any compensation of the negative impact on fecundity from hybrid dysgenesis
190 would be favoured. However, due to the pleiotropic effect of Nups on neuronal wiring and
191 behaviour, the escape from hybrid incompatibility could produce changes leading to
192 behavioural isolation (Munafo, et al. 2021; Nallasivan, et al. 2021). Since mutations in
193 promoters can exert a dominant effect (Pastinen, et al. 2004; Ronald, et al. 2005; Wittkopp, et
194 al. 2004), such changes could be rapidly fixed in contrast to recessive changes in coding
195 regions, which would remain hidden in heterozygosity. Our systematic analysis of the
196 evolution of all Nups coding and promoter regions suggests a less studied mode of evolution

197 through changes of sequences upstream of the transcription start site (TSS), particularly in the
198 promoters of core Nups, and supports the possibility that compensation of deleterious changes
199 in the germline can lead to altered neuronal wiring and rapid fixation of adaptive traits.
200

201 **Materials and Methods**

202 **Open Reading Frame Analysis**

203 To determine whether the ORF of nucleoporins is under selection, the PopFly online
204 database (imkt.uab.cat) developed from the Drosophila Genome Nexus project assembling
205 sequence data of around 1100 *Drosophila* genomes were used to analyse polymorphism data
206 from the ancestral Congo population because it is a sub-Saharan population with higher
207 ancestral stability than other populations (Lack, et al. 2015; Lack, et al. 2016; Mackay, et al.
208 2012). Nonsynonymous to synonymous polymorphisms within *D. melanogaster* or between *D.*
209 *melanogaster* and *D. simulans* were obtained and significance determined by Fisher's Exact
210 Test.
211

212 **Sequence Retrieval and Alignment**

213 Nucleoporin gene and promoter sequences were retrieved from UCSC genome browser
214 (genome.ucsc.edu) by entering the species and gene name, and then define the gene region
215 under tools/table browser. Alignments were done with clustalW within the MEGA program
216 (Kumar, et al. 2018). We also used MEGA to calculate mean overall nucleotide distance and
217 DnaSP (Rozas, et al. 2017) for the proportion of indels events in the alignments.
218

219 **Accumulation of Substitutions Along Extended Gene Regions**

220 To test for nonrandom accumulation of substitutions along the Nup extended gene
221 regions we determined significant deviations from a uniform distribution of substitutions using

222 an empirical cumulative distribution function. The function (G) detects monotonic increases in
223 substitutions (n) measured as the difference between the relative occurrence of a nucleotide
224 change and its relative position in the alignment (Civetta, et al. 2016). Whether differences
225 between the values of the G function (ΔG) between substitutional events deviates from a
226 random accumulation of changes is tested using Monte Carlo simulations to produce 100,000
227 samples of n events by sampling sites without replacement along the alignment (Civetta, et al.
228 2016).
229

230 **Comparison of Substitution Rates**

231 To analyse the promoters of Nups and the control group of the m6A writer complex
232 and readers a region of 1000 bp upstream and downstream of the transcriptional start site (TSS)
233 was used with TSS as an anchor point. Regions of differential alignment between the
234 *Drosophila* species were averaged across the NPC subcomplexes. A sliding window of five
235 bps upstream and downstream of each position was summed and processed in Excel to generate
236 heatmaps for highlighting hot spots of sequence divergence.
237

238 **Supplementary Material**

239 Supplementary Material is available at Genome Biology and Evolution online.
240

241 **Acknowledgements**

242 This work was supported by the Medical Research Council by D.W.J.M., the Biotechnology
243 and Biological Science Research Council to M.S. and the Natural Sciences and Engineering
244 Research Council of Canada to A.C. We thank R. Arnold for help with bioinformatics.
245

246 **Authors Contributions**

247 M.S. conceived and directed the project. All authors analysed data. M.S., A.C. and D.W.J.M.
248 wrote the manuscript. All authors read and approved the final manuscript.
249
250 **Data Availability**
251 All data generated or analysed during this study are included in Supplementary Dataset S1.
252
253 **Literature Cited**
254 Avila FW, Sirot LK, LaFlamme BA, Rubinstein CD, Wolfner MF 2011. Insect seminal fluid
255 proteins: identification and function. Annu Rev Entomol 56: 21-40. doi: 10.1146/annurev-
256 ento-120709-144823
257 Balacco DL, Soller M 2019. The m(6)A Writer: Rise of a Machine for Growing Tasks.
258 Biochemistry 58: 363-378. doi: 10.1021/acs.biochem.8b01166
259 Bawankar P, et al. 2021. Hakai is required for stabilization of core components of the m(6)A
260 mRNA methylation machinery. Nat Commun 12: 3778. doi: 10.1038/s41467-021-23892-5
261 Blumenstiel JP, Hartl DL 2005. Evidence for maternally transmitted small interfering RNA in
262 the repression of transposition in Drosophila virilis. Proc Natl Acad Sci U S A 102: 15965-
263 15970. doi: 10.1073/pnas.0508192102
264 Brennecke J, et al. 2008. An epigenetic role for maternally inherited piRNAs in transposon
265 silencing. Science 322: 1387-1392. doi: 10.1126/science.1165171
266 Cabal GG, et al. 2006. SAGA interacting factors confine sub-diffusion of transcribed genes to
267 the nuclear envelope. Nature 441: 770-773.
268 Chapman T 2018. Sexual conflict: mechanisms and emerging themes in resistance biology.
269 The American Naturalist 192: 217-229.

270 Civetta A 2016. Misregulation of Gene Expression and Sterility in Interspecies Hybrids:
271 Causal Links and Alternative Hypotheses. J Mol Evol 82: 176-182. doi: 10.1007/s00239-016-
272 9734-z
273 Civetta A, Ostapchuk DC, Nwali B 2016. Genome Hotspots for Nucleotide Substitutions and
274 the Evolution of Influenza A (H1N1) Human Strains. Genome Biol Evol 8: 986-993. doi:
275 10.1093/gbe/evw061
276 Coyne JA, Orr HA. 2004. Speciation. Sunderland, MA: Sinauer Associates.
277 Cutter AD, Bundus JD 2020. Speciation and the developmental alarm clock. Elife 9. doi:
278 10.7554/eLife.56276
279 Czech B, et al. 2018. piRNA-Guided Genome Defense: From Biogenesis to Silencing. Annu
280 Rev Genet 52: 131-157. doi: 10.1146/annurev-genet-120417-031441
281 Dezi V, Ivanov C, Haussmann IU, Soller M 2016. Nucleotide modifications in messenger
282 RNA and their role in development and disease. Biochem Soc Trans 44: 1385-1393. doi:
283 10.1042/bst20160110
284 Gillott C 2003. Male accessory glandd secretions: Modulators of Female Reproductive
285 Physiology and Behavior. Annu. Rev. Entomol. 48: 163-184.
286 Grayson P 2015. Izumo1 and Juno: the evolutionary origins and coevolution of essential
287 sperm-egg binding partners. Royal Society Open Science 2: 150296. doi:
288 doi:10.1098/rsos.150296
289 Hagen JFD, et al. 2019. tartan underlies the evolution of Drosophila male genital
290 morphology. Proceedings of the National Academy of Sciences of the United States of
291 America 116: 19025-19030. doi: 10.1073/pnas.1909829116
292 Hill MS, Vande Zande P, Wittkopp PJ 2021. Molecular and evolutionary processes
293 generating variation in gene expression. Nature Reviews Genetics 22: 203-215. doi:
294 10.1038/s41576-020-00304-w

295 Hollis B, et al. 2019a. Sexual conflict drives male manipulation of female postmating
296 responses in Drosophila melanogaster. Proc Natl Acad Sci U S A 116: 8437-8444. doi:
297 10.1073/pnas.1821386116
298 Hollis B, et al. 2019b. Sexual conflict drives male manipulation of female postmating
299 responses in Drosophila melanogaster. Proceedings of the National Academy of Sciences
300 116: 8437-8444.
301 Hoogendoorn B, et al. 2003. Functional analysis of human promoter polymorphisms. Human
302 Molecular Genetics 12: 2249-2254. doi: 10.1093/hmg/ddg246
303 Hopkins BR, Perry JC 2022. The evolution of sex peptide: sexual conflict, cooperation, and
304 coevolution. Biological Reviews. doi: 10.1111/brv.12849
305 King MC, Wilson AC 1975. Evolution at two levels in humans and chimpanzees. Science
306 188: 107-116. doi: 10.1126/science.1090005
307 Knuckles P, et al. 2018. Zc3h13/Flacc is required for adenosine methylation by bridging the
308 mRNA-binding factor Rbm15/Spenito to the m(6)A machinery component Wtap/Fl(2)d.
309 Genes Dev 32: 415-429. doi: 10.1101/gad.309146.117
310 Kubli E 2003. Sex-peptides: seminal peptides of the Drosophila male. Cell Mol Life Sci 60:
311 1689-1704.
312 Kumar S, Stecher G, Li M, Knyaz C, Tamura K 2018. MEGA X: Molecular Evolutionary
313 Genetics Analysis across Computing Platforms. Mol Biol Evol 35: 1547-1549. doi:
314 10.1093/molbev/msy096
315 Lack JB, et al. 2015. The Drosophila genome nexus: a population genomic resource of 623
316 Drosophila melanogaster genomes, including 197 from a single ancestral range population.
317 Genetics 199: 1229-1241. doi: 10.1534/genetics.115.174664

318 Lack JB, Lange JD, Tang AD, Corbett-Detig RB, Pool JE 2016. A Thousand Fly Genomes:
319 An Expanded Drosophila Genome Nexus. Mol Biol Evol 33: 3308-3313. doi:
320 10.1093/molbev/msw195
321 Lin DH, Hoelz A 2019. The Structure of the Nuclear Pore Complex (An Update). Annu Rev
322 Biochem 88: 18.11-18.59. doi: 10.1146/annurev-biochem-062917-011901
323 Mack KL, Nachman MW 2017. Gene Regulation and Speciation. Trends Genet 33: 68-80.
324 doi: 10.1016/j.tig.2016.11.003
325 Mackay TF, et al. 2012. The Drosophila melanogaster Genetic Reference Panel. Nature 482:
326 173-178. doi: 10.1038/nature10811
327 Malone CD, Hannon GJ 2009. Molecular evolution of piRNA and transposon control
328 pathways in Drosophila. Cold Spring Harb Symp Quant Biol 74: 225-234. doi:
329 10.1101/sqb.2009.74.052
330 Munafo M, et al. 2021. Channel nuclear pore complex subunits are required for transposon
331 silencing in Drosophila. Elife 10: e66321. doi: 10.7554/eLife.66321
332 Nallasivan MP, Haussmann IU, Civetta A, Soller M 2021. Channel nuclear pore protein 54
333 directs sexual differentiation and neuronal wiring of female reproductive behaviors in
334 Drosophila. BMC biology 19: 1-14.
335 Pascual-Garcia P, Capelson M 2014. Nuclear pores as versatile platforms for gene regulation.
336 Current opinion in genetics & development 25: 110-117.
337 Pastinen T, et al. 2004. A survey of genetic and epigenetic variation affecting human gene
338 expression. Physiological genomics.
339 Patlar B, Jayaswal V, Ranz JM, Civetta A 2021. Nonadaptive molecular evolution of seminal
340 fluid proteins in Drosophila. Evolution; International Journal of Organic Evolution 75: 2102 -
341 2113.

342 Presgraves DC 2010. The molecular evolutionary basis of species formation. Nature Reviews
343 Genetics 11: 175-180.
344 Presgraves DC, Balagopalan L, Abmayr SM, Orr HA 2003. Adaptive evolution drives
345 divergence of a hybrid inviability gene between two species of Drosophila. Nature 423: 715-
346 719.
347 Ronald J, Brem RB, Whittle J, Kruglyak L 2005. Local regulatory variation in
348 Saccharomyces cerevisiae. PLoS genetics 1: e25.
349 Rowe M, et al. 2020. Molecular Diversification of the Seminal Fluid Proteome in a Recently
350 Diverged Passerine Species Pair. Mol Biol Evol 37: 488-506. doi: 10.1093/molbev/msz235
351 Rozas J, et al. 2017. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets.
352 Mol Biol Evol 34: 3299-3302. doi: 10.1093/molbev/msx248
353 Seeholzer LF, Seppo M, Stern DL, Ruta V 2018. Evolution of a central neural circuit
354 underlies Drosophila mate preferences. Nature 559: 564-569. doi: 10.1038/s41586-018-0322-
355 9
356 Sirot LK, Wong A, Chapman T, Wolfner MF 2014. Sexual conflict and seminal fluid
357 proteins: a dynamic landscape of sexual interactions. Cold Spring Harb Perspect Biol 7:
358 a017533. doi: 10.1101/cshperspect.a017533
359 Smale ST, Kadonaga JT 2003. The RNA polymerase II core promoter. Annu Rev Biochem
360 72: 449-479. doi: 10.1146/annurev.biochem.72.121801.161520
361 Soller M, Bownes M, Kubli E 1999. Control of oocyte maturation in sexually mature
362 Drosophila females. Dev Biol 208: 337-351.
363 Stern DL 2000. Perspective: evolutionary developmental biology and the problem of
364 variation. Evolution 54: 1079-1091.
365 Taddei A, et al. 2006. Nuclear pore association confers optimal expression levels for an
366 inducible yeast gene. Nature 441: 774-778.

367 Tang S, Presgraves DC 2009. Evolution of the Drosophila nuclear pore complex results in
368 multiple hybrid incompatibilities. Science 323: 779-782.
369 Tang S, Presgraves DC 2015. Lineage-specific evolution of the complex Nup160 hybrid
370 incompatibility between Drosophila melanogaster and its sister species. Genetics 200: 1245-
371 1254.
372 Taylor MS, et al. 2006. Heterotachy in mammalian promoter evolution. PLoS genetics 2: e30.
373 Vo Ngoc L, Kassavetis GA, Kadonaga JT 2019. The RNA Polymerase II Core Promoter in
374 Drosophila. Genetics 212: 13-24. doi: 10.1534/genetics.119.302021
375 Wilkins AS. 2002. The evolution of developmental pathways: Sinauer Associates.
376 Wittkopp PJ, Haerum BK, Clark AG 2004. Evolutionary changes in cis and trans gene
377 regulation. Nature 430: 85-88.
378 Wray GA, et al. 2003. The evolution of transcriptional regulation in eukaryotes. Mol Biol
379 Evol 20: 1377-1419. doi: 10.1093/molbev/msg140
380

381 **Figure Legends**

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

384 **A**) Graphic depiction of the nuclear pore complex with differentially coloured subcomplexes
385 inner ring (IR, pink), outer ring (OR, blue), cytoplasmic filament (CF, orange), nuclear basket
386 (yellow) and pore membrane proteins (POMS, red) embedded in the nuclear envelope lipid-
387 bilayer (green).

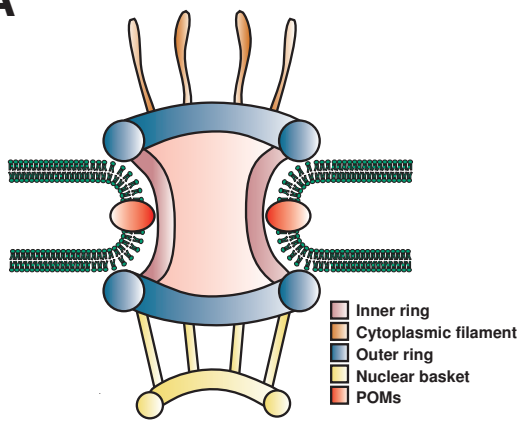
388 **B)** Significance of synonymous to non-synonymous changes in open reading frames (ORFs)
389 indicated by an asterisk and p-values.
390

391 **Fig. 2: Promoter regions of channel Nups 58 and 62 have diverged in closely related**
392 **species.**
393 **A, B)** Sequence alignment of the Nup58 and Nup62 promoter regions from closely related
394 species. Nucleic acids changes from *D. melanogaster* are indicated in black. Transcribed parts
395 of the *Vha13* and *CG7997* 3′UTR and the *Nup58* and *Nup62* 5′UTR are indicated by a line.
396 **C, D)** Plot of cumulative differences along the sequence (G) between the relative occurrences
397 of indels and their position from the alignment of the gene region around *Nup58* and *Nup62*
398 between *D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba* and *D. erecta*. Positions in the
399 alignment with significant stretches of substitutions are identified by black line(s).
400
401 **Fig. 3: Core nuclear pore protein genes rapidly evolve through indel accumulation in**
402 **promoters.**
403 **A)** Heatmaps indicating divergence for indels in yellow among closely related *D.*
404 *melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba* and *D. erecta* for the control group of genes
405 (m6A writer complex and readers) and the inner ring (IR), outer ring (OR), cytoplasmic
406 filament (CF) and nuclear basket (NB) below the gene model with the transcription start site
407 (TSS) indicated by an arrow. The red line indicates the promoter region used for quantification
408 of the substitution rate in D.
409 **B)** Quantification of the change rate of indels in the TATA box distal region (from -30 to -
410 380). Statistically significant differences from non-parametric chi-squared tests are indicated
411 by asterisks (* $p \leq 0.05$, ** $p \leq 0.001$, *** $p \leq 0.0001$).
412
413 **Supplementary fig. S1: Comparison of sequence changes in gene upstream and**
414 **downstream regions.**

415 **A)** Gene model depicting the intergenic region followed by the transcription start site (TSS)
416 and the genespan with upstream (US) and downstream (DS) regions analysed separately.
417 **B)** Analysis of nucleotide changes and indels within closely related species of Drosophila (*D.*
418 *melanogaster, D. simulans, D. sechellia, D. yakuba and D. erecta*) through comparison of the
419 proportion of indel events and nucleotide distance. Statistically significant differences between
420 the upstream and downstream regions were calculated from unpaired t-tests and indicated by
421 asterisks (* $p \leq 0.05$, ** $p \leq 0.001$, *** $p \leq 0.0001$).
422
423 **Supplementary fig. S2: Inner and outer ring nuclear pore protein gene subcomplexes**
424 **undergo rapid evolution through indel accumulation in promoters.**
425 **A)** Gene model depicting the intergenic region followed by the transcription start site (TSS)
426 and the genespan with the 350 bp region analysed for promotor activity upstream of the TATA
427 box location.
428 **A-H)** Heatmaps indicating divergence in yellow among closely related Drosophila species for
429 the control group of genes (m6A writer complex and readers) and the inner ring (IR), outer ring
430 (OR), cytoplasmic filament (CF) and nuclear basket (NB). Calculated sequence difference
431 index, deletion index, insertion index and base change index events are reported as bar charts
432 for the NPC subgroups compared to the control group. Index events were defined by values
433 above the control intergenic region average. Statistically significant differences between the
434 NPC subgroups and the control group were calculated from chi-squared tests and indicated by
435 asterisks (* $p \leq 0.05$, ** $p \leq 0.001$, *** $p \leq 0.0001$).
436

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

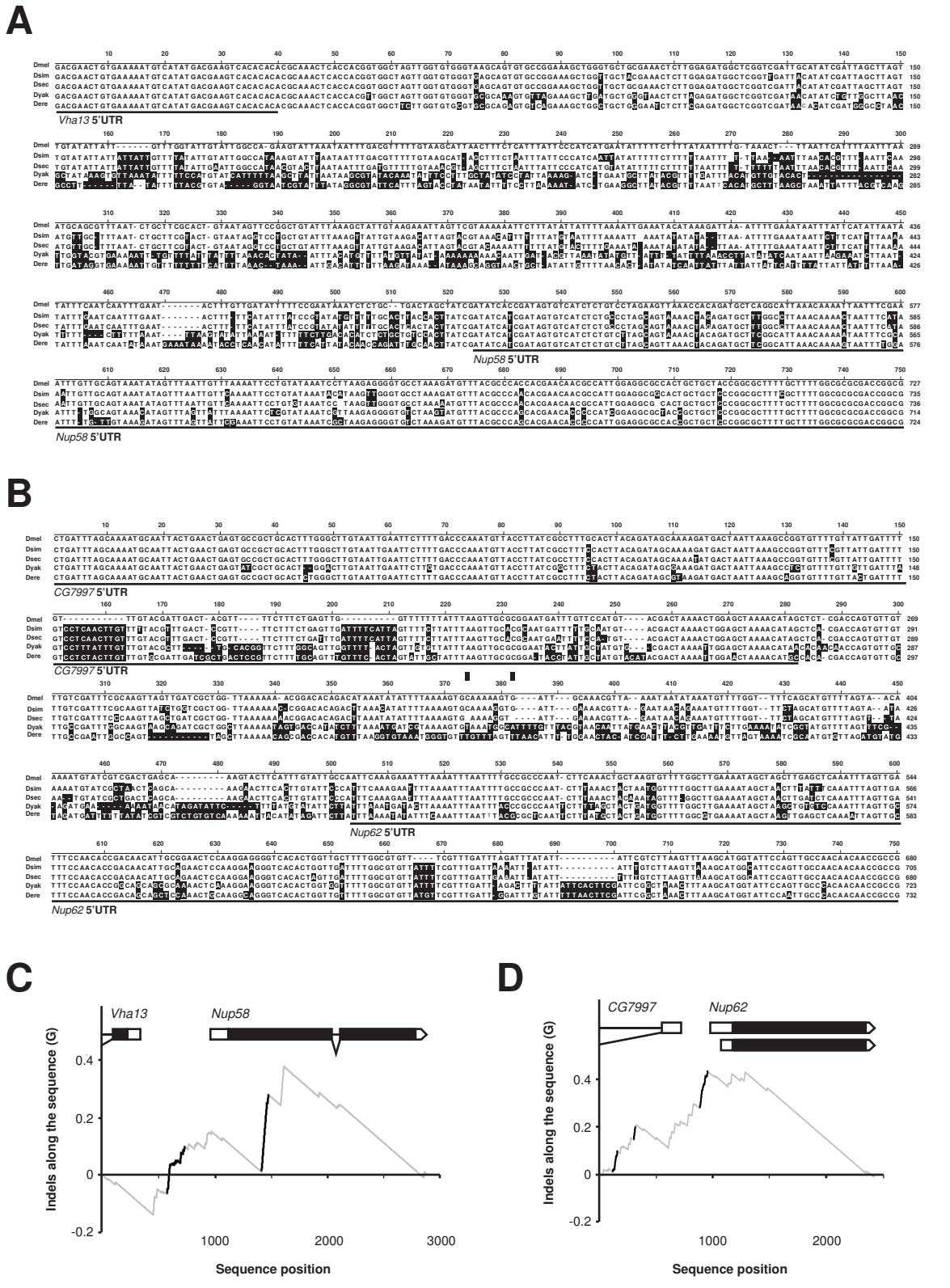
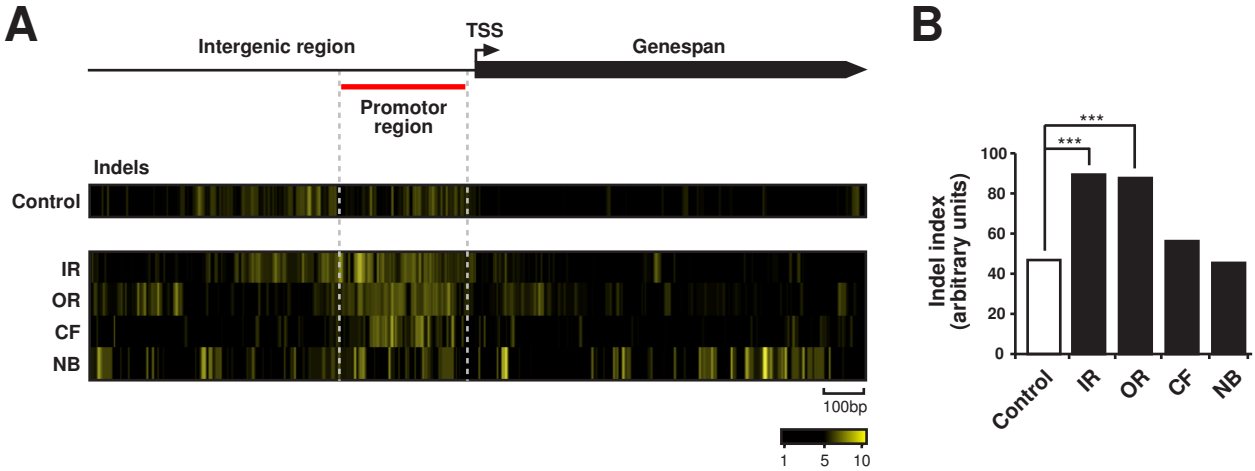
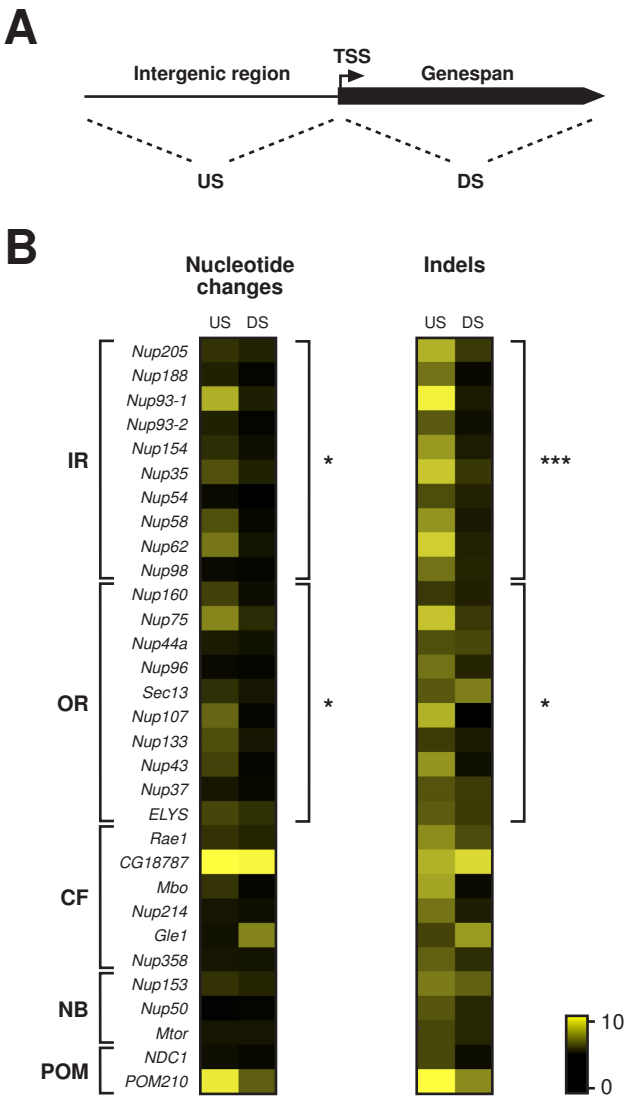
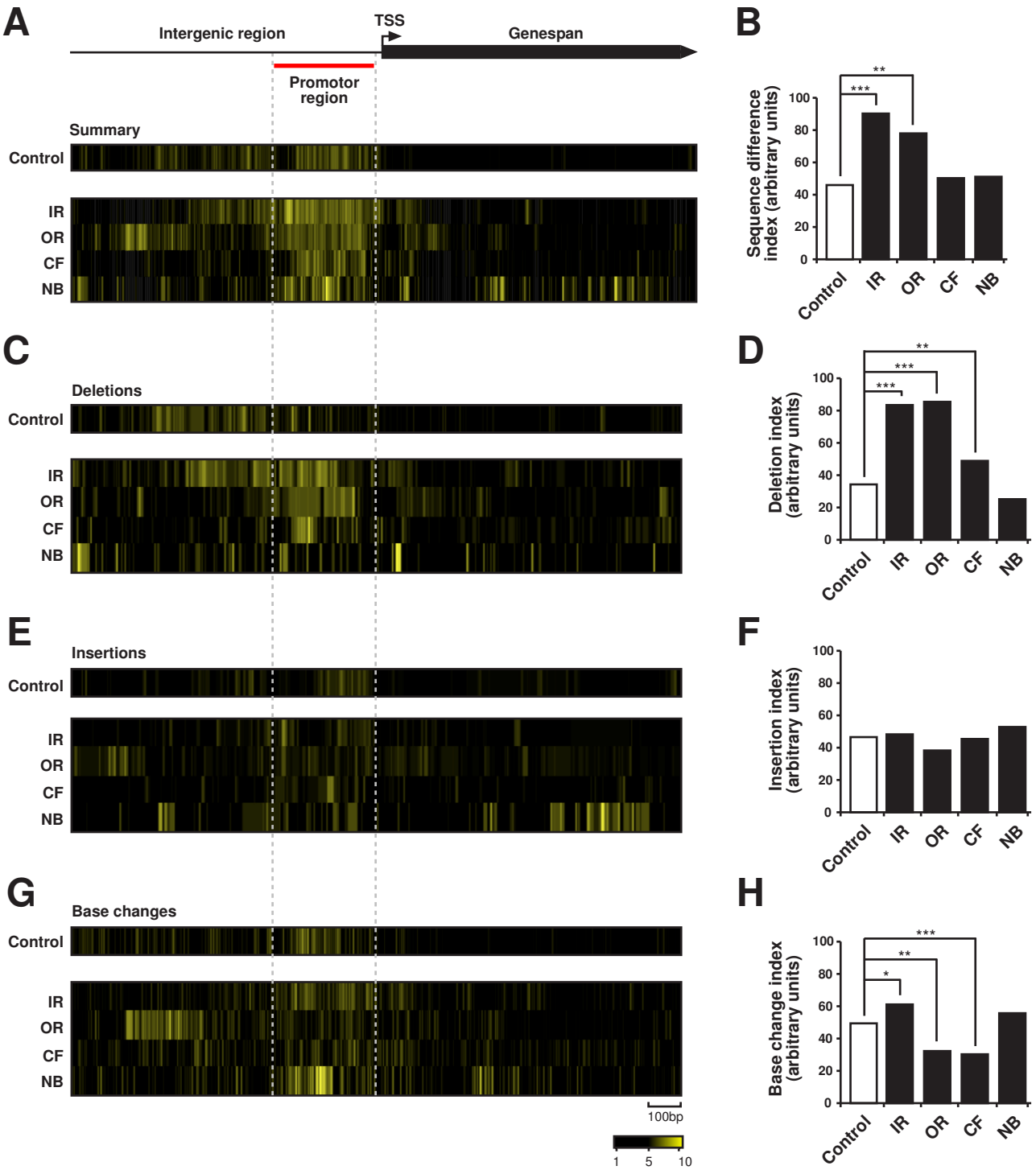

A
Dmel
Dsim
Dsec
Dyak
Dere
Vha13 5'UTR
Nup58 5'UTR
B
CG7997 5'UTR
Nup62 5'UTR
C
Vha13
Nup58
Indels along the sequence (G)
Sequence position
D
CG7997
Nup62
Indels along the sequence (G)
Sequence position

Fig. 3


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

Supplementary fig. S1



Supplementary fig. S2