Very small spontaneous structural mutation rate in green algae

Matéo Léger-Pigout[1], Marc Krasovec[1*]

[1] CNRS, Biologie Intégrative des Organismes Marins (BIOM), Observatoire Océanologique, Banyuls sur mer, 66650, France

* marc.krasovec@obs-banyuls.fr

RUNNING TITLE

Structural mutation rate in green algae

KEY WORDS

mutation rate, structural mutation, mutation accumulation, phytoplankton

ABSTRACT

Structural mutations are very important in evolution and led to major innovations, but our knowledge of the spontaneous structural mutation rate is very limited. We so cannot have a complete view of the adaptive potential of species and new variant in a population without addressing this lack. Here, we used Illumina genomic data of mutation accumulation experiments from five green algae species and measured very low structural mutation rate, between $5x10^{-12}$ to $2x10^{-11}$ mutations per site, at least ten times lower than the nucleotide mutation rate in these species. Such low rates of structural mutations are typical of species with large effective population sizes, and suggests a very strong selection toward a small structural mutation rate. Last, we show that the rate of true positive detection of de novo structural mutations is variant types dependent and globally poor, highlighting the necessity of long read sequencing in future mutation accumulation studies.

INTRODUCTION

Spontaneous mutations renew the standing genetic variation and permit the adaptation of species through natural selection. The spontaneous mutation rate μ is so one of the most important parameters to know in evolution. Since decades biologists tried to measure it using mutation accumulation (MA) experiment (Halligan & Keightley, 2009), indirectly with a phenotypic approach by following fitness variation of MA lines (Mukai, 1964), and then directly by sequencing (Lynch *et al.*, 2016; Katju & Bergthorsson, 2019). Our knowledge made significant progresses the last fifteen years, particularly on the nucleotide mutation rate μ_{bs} (Lynch *et al.*, 2016; Katju & Bergthorsson, 2019) and short insertion-deletion (indel) mutation rate μ_{id} (Sung *et al.*, 2016). We observe a large variation of μ_{bs} and μ_{id} between species, from ~$1x10^{-11}$ to ~$1x10^{-8}$ in ciliates and humans, respectively. This variation is in part explained by the drift barrier hypothesis (Sung *et al.*, 2012) where high effective population size (N_e) species reaches lower μ than species with small N_e. This is due to the strength of selection that pushes harder toward a small μ (and so limiting the mutation load) in large population where drift is limited. This hypothesis is the most accepted to explain the mutation rate variation between species, even if other factors impact it but with smaller effect such as GC content (Krasovec *et al.*, 2017), genome size, generation time or metabolic rate (Martin & Palumbi, 1993; Mooers & Harvey, 1994; Thomas *et al.*, 2010; Weller & Wu, 2015). Both μ_{bs} and μ_{id} depends so mostly of N_e, but our knowledge about the spontaneous structural mutation rate μ_{st} is very poor (Press *et al.*, 2019; Ho & Schaack, 2021) and available in only few models (see Table 3 of Katju and Bergthorsson 2019 for copy number variant for example). Structural mutations, defined here as duplications, large insertions-deletions, inversions or chromosome rearrangements, may have stronger phenotypic effect than nucleotide and short indels mutations because they impact a larger portion of the genome. Thus, the pressure toward a small μ_{st} should be stronger than for μ_{bs} and μ_{id}. In addition, it is impossible to have a complete few of new variants and adaptive potential of a species with limited knowledges of μ_{st}. Structural mutations led to major evolutive innovations, for example the supergenes following inversions. In birds, this process saw the emergence of different morphotypes in a same species with different colours, behaviours and feather shapes (Küpper *et al.*, 2016; Tuttle *et al.*, 2016). Here, we measured μ_{st} for five species of green algae using mutation accumulation experiment Illumina reads from previous publications (Krasovec *et al.*, 2016, 2017, 2018).

MATERIALS AND METHODES

Raw reads (150 and 300 bp paired-end) were mapped to the reference genomes with bwa mem (Li & Durbin, 2010), then bam files were treated with samtools (Li *et al.*, 2009) and structural variant calling was done with delly (Rausch *et al.*, 2012), svaba (Wala *et al.*, 2018) and lumpy (Layer *et al.*, 2014) with standard parameters (excepted ploidy at 1). For svaba, we used the option -n to input the T_0 genome bam file as control. VCF files were then sorted with several threshold to increase the rate of true de novo mutation identification: the mutated site must be covered by 10 in the concerned MA line and the T0 genome; the mapping quality of the site must be higher than 20; the variant must be present in only one MA line, expected for few variants present in two or three MA lines because of non-independent lines (in that case common mutations and generations were count as one); the variant must totalize all the coverage of the site; last, a part of the structural variants appeared in only one MA line, but were not count as a mutation candidate because other lines had same variants but at different positions within the confidence interval of start-end positions. De novo structural mutations candidates were then checked by PCR and Sanger sequencing when possible and by coverage with samtools depth.

RESULTS AND DISCUSSION

False - true positive rate of structural de novo mutation detection

The total number of de novo structural mutations candidates was 26 considering the 5 species and the 3 software (Table 1), but PCR and Sanger sequencing check considerably reduced this number to 1, given a very low true positive rate of 4%. From Sanger sequencing, the 8 tested breaks from svaba and delly were false, so we considered all other non-tested breaks (10) as false. The other variants were deletions (5) and duplications (3), all larger than 300 nt apart one deletion of 215 pb. According to Wala et al (Wala *et al.*, 2018), the positive rates of deletions larger than 300 bp are 85% for svaba and delly, and 95% for lumpy; but for insertion/duplication, the validation rates decrease to 37%, 24% and 29%. Here, the positive rate of deletion is 20%, substantially lower than previously reported. Regarding duplications, the 3 candidates were false according to PCR and coverage check (two candidates were not verified by PCR). The true positive rate of duplications is so 0% here, but with only 3 candidates this rate should be taken with caution. Overall, these results give an idea of the difficulties to identify structural de novo mutations with good accuracy particularly with short reads. This result also shows that structural mutations are very rare in these types of experiments, making any statistics on structural mutations spectrum and distribution not often possible. In yeast, regarding duplications, only three events were identified in a previous study with 145 lines maintained for about 2,000 generation each (Zhu *et al.*, 2014).

μ_{st} is more than 10 time lower than μ_{bs} and μ_{id}

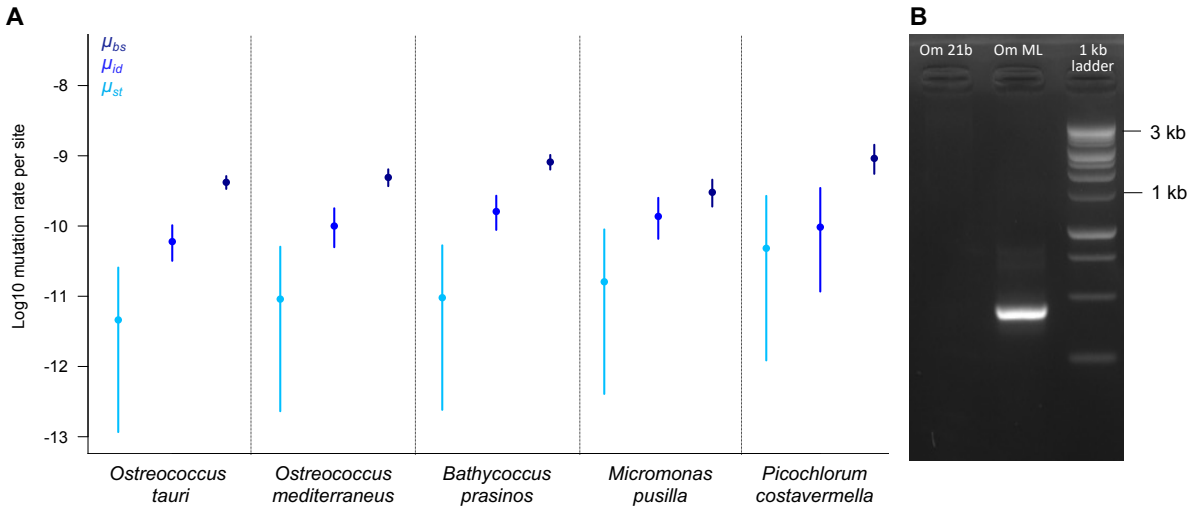
The single de novo mutation validated by PCR is a deletion of 1,382 bp identified by lumpy in *Ostreoccocus mediterraneus* on a very variable chromosome in this species, named outlier chromosome (Yau *et al.*, 2020). The mutated region is intergenic and the deleted sequence appears elsewhere in the genome (Table S1). This makes this deleted sequence a possible repeat region resulting from transposable elements, commonly known to be implicated if this type of mutation. This single mutation gives a very low μ_{st} for the 5 species, between $5x10^{-12}$ and $2x10^{-11}$ structural mutation per site per haploid genome per generation (Figure 1). It is at least ten times lower than μ_{bs} and μ_{id} in these species. Taking all species together, $\mu_{st}=2x10^{-12}$. μ_{st} in green algae are so among the lower known μ in the tree of life, similar to μ_{bs} in *Paramecium* (Long *et al.*, 2018) and *Dictostelium* (Kucukyildirim *et al.*, 2020). The selection pressure toward a low μ in *Paramecium* species is linked to their very high N_e of ~ 100 million (Sung *et al.*, 2012). In phytoplankton species, only few N_e estimations are available but all turn around few millions to 10 million (Blanc-Mathieu *et al.*, 2017; Krasovec *et al.*, 2019, 2020). μ_{bs} and μ_{id} in phytoplankton are at least ten time higher than in *Paramecium*. This could indicate that the selection pressure on μ_{st} in the green algae is very strong and reduces μ_{st} to a value that only very high N_e species such as *Paramecium* can reach for μ_{bs}. One of the origins of such strong selection may be the genome structure of green algae studied here. Their genomes are small, very compacted with few introns and small distance between genes. For example, *O. tauri* has a genome size of 12.5 Mb, with 81.6% of coding sequences, 82% of genes without introns and an average distance between genes of 254 bp. In that case, any structural mutation has high chance to touch at least one gene or its regulatory region, making almost all structural mutations deleterious. We can hypothesis that it is less the case in larger genomes with a longer distance between genes and high proportion of repeats and non-coding sequences. Another explanation of the low structural mutation rate may come from the mutation mechanisms themself. Many mechanisms act directly on nucleotides such as methylation and oxidation and are known to increase μ_{bs}, making it higher than μ_{st} or μ_{id} just because single sites have more chance to mutate. Last explanation may be a method bias, *e.g* the short illumina reads are not the best strategy to detect structural variants. Further researches with mutation accumulation experiments coupled with long reads sequencing is the next milestone significantly increase our knowledge on μ_{st}.

147 **Table 1.** List of de novo structural mutation candidates. PCR-Sanger and Covergae are the two checking
148 method for de novo mutations. Only the deletion of Om21b is true.

MA line	Mutation type	Software	Start	End	PCR - Sanger	Coverage
Ot20	break	delly	NC_014444.2: 184889	NC_014431.2: 783883	false	na
Om14	break	delly	BCC102-800924: 233	BCC102-1186077: 1181287	false	na
Om17 - 18	break	delly	BCC102-594986: 60163	BCC102-768578: 224081	false	na
Om1	break	delly	BCC102_ML_SOC: 68272	BCC102-397956: 160264	false	na
Om2	break	delly	BCC102-466838: 2048	BCC102-867403: 392	false	false
Ot38	deletion	delly	NC_014427.2: 430027	NC_014427.2: 438281	no check	false
Bp8	deletion	delly	NC_023992.1: 297114	NC_023992.1: 301342	false	false
Bp20	duplication	delly	NC_023992.1: 296522	NC_023992.1: 297150	false	false
Ot38	break	svaba	NC_014427.2: 820183	NC_014428.2:994946	false	na
Ot38	break	svaba	NC_014433.2: 406614	NC_014437.2: 310849	false	na
Om13	break	svaba	BCC102-873724: 213431	BCC102_ML_SOC: 100063	false	na
Ot22	break	lumpy	NC_014426.2: 181294	NC_014428: 229429	no check	na
Ot23	break	lumpy	NC_014426.2: 929056	NC_014442.2: 399570	no check	na
Ot9 - 10	break	lumpy	NC_014435.2: 273589	NC_014439.2: 385664	no check	na
Ot20 - 31	break	lumpy	NC_014441.2: 17675	NC_014443.2: 58813	no check	na
Om38	break	lumpy	BCC102-993683: 535232	BCC102-768578: 224085	no check	na
Om12	break	lumpy	BCC102-637367: 375185	BCC102-637367: 319399	no check	na
Om3	break	lumpy	BCC102-873724: 329213	BCC102-768578: 224085	no check	na
Om13	break	lumpy	BCC102-873724: 173257	BCC102-546026: 350305	no check	na
Om28	break	lumpy	BCC102-867403: 378390	BCC102-496340: 275198	no check	na
Mp27 - 34 - 35	break	lumpy	chrom_01: 564769	chrom_01: 1360407	no check	na
Om9 - 29d	deletion	lumpy	BCC102-867403: 33955	BCC102-867403: 34170	false	false
Om31	deletion	lumpy	BCC102-466838: 274207	BCC102-466838: 276502	false	false
Om21b	deletion	lumpy	BCC102_ML_SOC: 571078	BCC102_ML_SOC: 572460	true	true
Om17	duplication	lumpy	BCC102-806365: 187888	BCC102-806365: 188243	no check	false
Om32	duplication	lumpy	BCC102_ML_SOC: 243207	BCC102_ML_SOC: 251990	no check	false

149

150



151

152 **Figure 1. A.** Spontaneous mutation rate per nucleotide per haploid genome per generation. Black: μ_{st},
153 spontaneous structural mutation rate. Dark blue: μ_{id}, spontaneous short insertion-deletion mutation rate.
154 Blue: μ_{bs}, spontaneous nucleotide mutation rate. Horizontal bars are confidence interval following a
155 Poisson distribution. μ_{id} and μ_{bs} come from previous studies (Krasovec *et al.*, 2017, 2018). μ_{st} for *O.*
156 *tauri*, *M. pusilla*, *B. prasinos* et *P. costavermella* are calculated assuming one hypothetical mutation (non
157 has been validated), so the values provided here are the above estimations. **B.** PCR product migration
158 of one large deletion (1,382 bp in intergenic region) identified in *O. mediterraneus*. The ladder is the 1
159 kb DNA ladder from Promega. No amplification is observed in the mutation accumulation line Om 21b
160 because primers were designed insight the deleted sequence. Om ML is the mother line of the mutation
161 accumulation (T0 genome).

DATA AVAILABILITY STATEMENT

All genomic raw reads used in this study are available on the SRA database in the NCBI under the bioprojects PRJNA531882 (*Ostreoccocus tauri*, *O. mediterraneus*, *Micromonas pusilla*, *Bathycoccus prasinos*), PRJNA453760 and PRJNA389600 (*Picochlorum costavermella*).

ACKNOWLEDGEMENTS

We are grateful to GENOPHY team of the Banyuls sur mer lab for discussions and advices.

AUTHOR CONTRIBUTIONS

MK designed and managed the study, MLP did bioinformatics analysis and wet lab works, MK wrote the draft of the manuscript, all authors contributed to editing the last manuscript version.

FUNDING

This work was funded by ANRJCJC-SVSE6-2013-0005.

COMPETING INTERESTS

The authors declare no conflicts of interest.

BIBLIOGRAPHY

Blanc-Mathieu, R., Krasovec, M., Hebrard, M., Yau, S., Desgranges, E., Martin, J., *et al.* 2017. Population genomics of picophytoplankton unveils novel chromosome hypervariability. *Sci. Adv.* **3**: e1700239.

Halligan, D.L. & Keightley, P.D. 2009. Spontaneous Mutation Accumulation Studies in Evolutionary Genetics. *Annu. Rev. Ecol. Evol. Syst.* **40**: 151–172.

Ho, E.K.H. & Schaack, S. 2021. Intraspecific variation in the rates of mutations causing structural variation in Daphnia magna. *Genome Biol. Evol.* evab241.

Katju, V. & Bergthorsson, U. 2019. Old Trade, New Tricks: Insights into the Spontaneous Mutation Process from the Partnering of Classical Mutation Accumulation Experiments with High-Throughput Genomic Approaches. *Genome Biol. Evol.* **11**: 136–165.

Krasovec, M., Eyre-Walker, A., Grimsley, N., Salmeron, C., Pecqueur, D., Piganeau, G., *et al.* 2016. Fitness Effects of Spontaneous Mutations in Picoeukaryotic Marine Green Algae. *G3 GenesGenomesGenetics* **6**: 2063–2071.

Krasovec, M., Eyre-Walker, A., Sanchez-Ferandin, S. & Piganeau, G. 2017. Spontaneous Mutation Rate in the Smallest Photosynthetic Eukaryotes. *Mol. Biol. Evol.* **34**: 1770–1779.

Krasovec, M., Rickaby, R.E.M. & Filatov, D.A. 2020. Evolution of Mutation Rate in Astronomically Large Phytoplankton Populations. *Genome Biol. Evol.* **12**: 1051–1059.

Krasovec, M., Sanchez-Brosseau, S., Grimsley, N. & Piganeau, G. 2018. Spontaneous mutation rate as a source of diversity for improving desirable traits in cultured microalgae. *Algal Res.* **35**: 85–90.

Krasovec, M., Sanchez-Brosseau, S. & Piganeau, G. 2019. First Estimation of the Spontaneous Mutation Rate in Diatoms. *Genome Biol. Evol.* **11**: 1829–1837.

Kucukyildirim, S., Behringer, M., Sung, W., Brock, D.A., Doak, T.G., Mergen, H., *et al.* 2020. Low Base-Substitution Mutation Rate but High Rate of Slippage Mutations in the Sequence

Repeat-Rich Genome of Dictyostelium discoideum. *G3 Bethesda Md* **10**: 3445–3452.
Küpper, C., Stocks, M., Risse, J.E., dos Remedios, N., Farrell, L.L., McRae, S.B., *et al.* 2016. A supergene determines highly divergent male reproductive morphs in the ruff. *Nat. Genet.* **48**: 79–83. Nature Publishing Group.
Layer, R.M., Chiang, C., Quinlan, A.R. & Hall, I.M. 2014. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol.* **15**: R84.
Li, H. & Durbin, R. 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**: 589–595.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al.* 2009. The Sequence Alignment/Map format and SAMtools. *Bioinforma. Oxf. Engl.* **25**: 2078–2079.
Long, H., Doak, T.G. & Lynch, M. 2018. Limited Mutation-Rate Variation Within the Paramecium aurelia Species Complex. *G3 GenesGenomesGenetics* **8**: 2523–2526.
Lynch, M., Ackerman, M.S., Gout, J.-F., Long, H., Sung, W., Thomas, W.K., *et al.* 2016. Genetic drift, selection and the evolution of the mutation rate. *Nat. Rev. Genet.* **17**: 704–714.
Martin, A.P. & Palumbi, S.R. 1993. Body size, metabolic rate, generation time, and the molecular clock. *Proc. Natl. Acad. Sci. U. S. A.* **90**: 4087–4091.
Mooers, A.O. & Harvey, P.H. 1994. Metabolic rate, generation time, and the rate of molecular evolution in birds. *Mol. Phylogenet. Evol.* **3**: 344–350.
Mukai, T. 1964. The Genetic Structure of Natural Populations of Drosophila Melanogaster. I. Spontaneous Mutation Rate of Polygenes Controlling Viability. *Genetics* **50**: 1–19.
Press, M.O., Hall, A.N., Morton, E.A. & Queitsch, C. 2019. Substitutions Are Boring: Some Arguments about Parallel Mutations and High Mutation Rates. *Trends Genet. TIG* **35**: 253–264.
Rausch, T., Zichner, T., Schlattl, A., Stütz, A.M., Benes, V. & Korbel, J.O. 2012. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**: i333–i339.
Sung, W., Ackerman, M.S., Dillon, M.M., Platt, T.G., Fuqua, C., Cooper, V.S., *et al.* 2016. Evolution of the Insertion-Deletion Mutation Rate Across the Tree of Life. *G3 Bethesda Md*, doi: 10.1534/g3.116.030890.
Sung, W., Ackerman, M.S., Miller, S.F., Doak, T.G. & Lynch, M. 2012. Drift-barrier hypothesis and mutation-rate evolution. *Proc. Natl. Acad. Sci. U. S. A.* **109**: 18488–18492.
Thomas, J.A., Welch, J.J., Lanfear, R. & Bromham, L. 2010. A generation time effect on the rate of molecular evolution in invertebrates. *Mol. Biol. Evol.* **27**: 1173–1180.
Tuttle, E.M., Bergland, A.O., Korody, M.L., Brewer, M.S., Newhouse, D.J., Minx, P., *et al.* 2016. Divergence and Functional Degradation of a Sex Chromosome-like Supergene. *Curr. Biol. CB* **26**: 344–350.
Wala, J.A., Bandopadhayay, P., Greenwald, N.F., O'Rourke, R., Sharpe, T., Stewart, C., *et al.* 2018. SvABA: genome-wide detection of structural variants and indels by local assembly. *Genome Res.* **28**: 581–591.
Weller, C. & Wu, M. 2015. A generation-time effect on the rate of molecular evolution in bacteria. *Evol. Int. J. Org. Evol.* **69**: 643–652.
Yau, S., Krasovec, M., Benites, L.F., Rombauts, S., Groussin, M., Vancaester, E., *et al.* 2020. Virus-host coexistence in phytoplankton through the genomic lens. *Sci. Adv.* **6**: eaay2587. American Association for the Advancement of Science.
Zhu, Y.O., Siegal, M.L., Hall, D.W. & Petrov, D.A. 2014. Precise estimates of mutation rate and spectrum in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **111**: E2310-2318.

258 **Table S1.** Blastn results of the deleted sequence in *Ostrecoccus mediterraneus*. The blasted sequence
259 was the sequence amplified by PCR in the MA mother line (404 bp length).

Chromosome	Identity	Length	Start	End	Pvalue
BCC102-546026	98.916	369	358951	359319	0.0
BCC102-546026	93.590	156	356541	356696	3.17e-62
BCC102-546026	89.062	128	468750	468877	5.46e-40
BCC102-466838	91.617	167	435797	435963	9.76e-62
BCC102_ML_SOC	99.010	404	321019	321422	0.0
BCC102_ML_SOC	98.721	391	571052	571442	0.0
BCC102_ML_SOC	90.678	354	225814	226166	2.59e-133
BCC102_ML_SOC	90.571	350	257243	257591	4.34e-131
BCC102_ML_SOC	90.571	350	584735	585083	4.34e-131
BCC102_ML_SOC	90.299	134	208824	208955	2.28e-44

260
261
262