

>> SEARCH FOR GENOMIC REARRANGEMENTS IN THE *POLYPEDILUM VANDERPLANKI* CELL LINE IN AN EXPERIMENT OF DESICCATION-REHYDRATION CYCLES

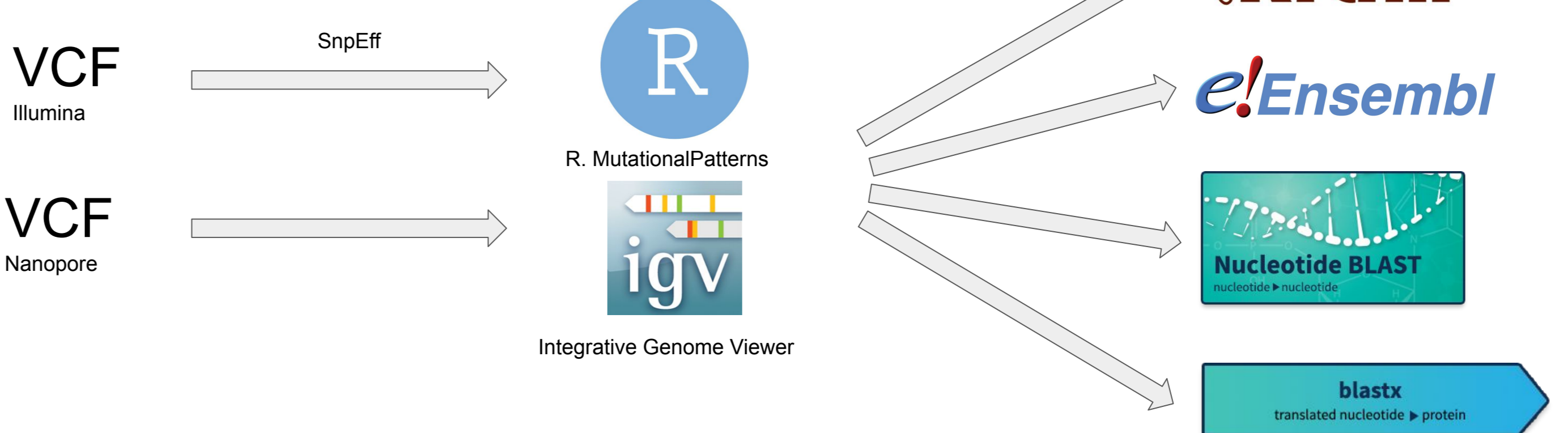
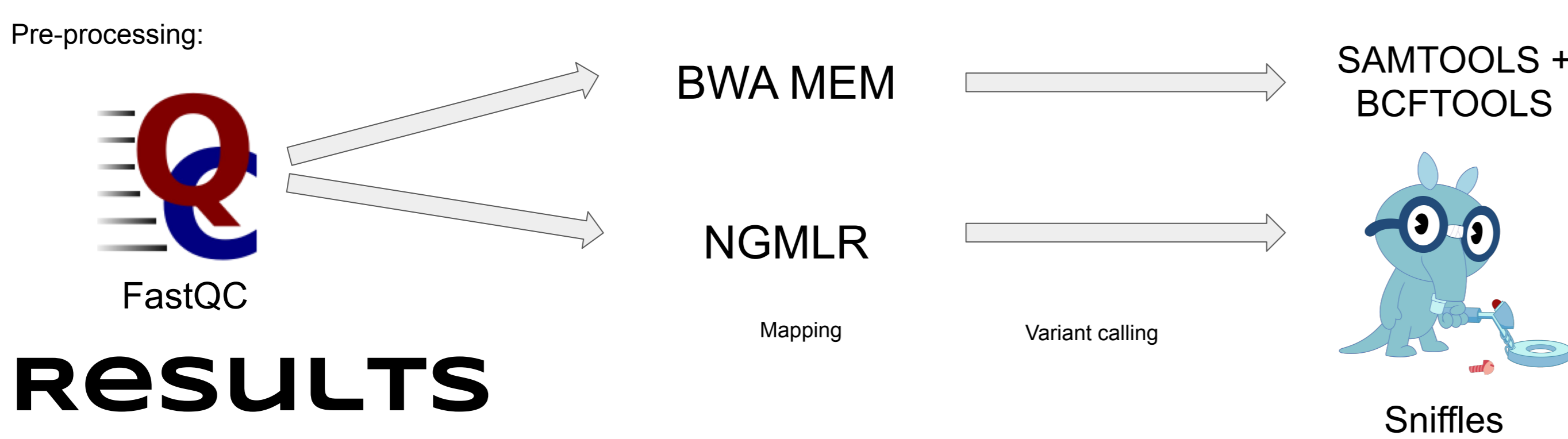
INTRODUCTION

Личиночная стадия экстремофильного комара-звонца *Polypedilum vanderplanki* является наиболее сложным организмом, способным к ангидробиозу (личинка теряет больше 97% всей воды в теле, замещая её трегалозой). При добавлении воды личинка может регидратироваться менее чем за час и вернуться к активной жизнедеятельности. В ходе эксперимента по изучению влияния сушки на клетки личинки было показано, что геномная ДНК фрагментируется за счет двухцепочечных разрывов (DSBs), но, что удивительно, за несколько дней ДНК восстанавливается. Мы предположили, что подобная фрагментация может привести к геномным перестройкам, а также к точечным мутациям на краях разрывов. Чтобы это проверить мы получили данные секвенирования генома клеточной линии Pv11, прошедшей цикл сушки-регидратации.

The larval stage of the extremophile midge *Polypedilum vanderplanki* is the most complex organism that is capable of anhydrobiosis (can survive the loss of about 97% of its body water) and rehydrates within less than an hour. Experiments with drying of a larval cell line Pv11 show that the larval genomic DNA is fragmented by double-strand breaks (DSBs), but, surprisingly, DNA is restored in a few days. We expect such fragmentation would lead to genomic rearrangements and point mutations. To verify this, we have analyzed sequence data of the Pv11 cells which had survived the desiccation-rehydration cycle.



METHODS



RESULTS

1. POINT MUTATIONS DISTRIBUTIONS PROBABLY NOT CHANGE IN EXPERIMENT

We have analyzed mutation patterns after filtering common substitutions between all the cells. No significant difference was found between the number of SNPs, types of base substitutions, and functional regions harboring substitutions.



Fig. 1. Algorithm of SNPs association with genomic rearrangements
Рис. 1. Алгоритм поиска SNP, ассоциированных с геномными перестройками

Table 1. SNPs associated with genomic rearrangements
Таблица 1. SNP, ассоциированных с перестройками

	Merged Before DR (two filtration)	Merged After DR Unique	Merged After DR (two filtration)	Merged After DR Unique
Frequency in region (-20...+20)	0.02335671	0.001168721	0.02228665	0.0007011958
Frequency in rest part	0.00666309	0.002899767	0.00653780	0.0001550589
SNP associated with SV	1319	66	2352	74
Ratio	3.50539	4.03095	3.40891	4.522125

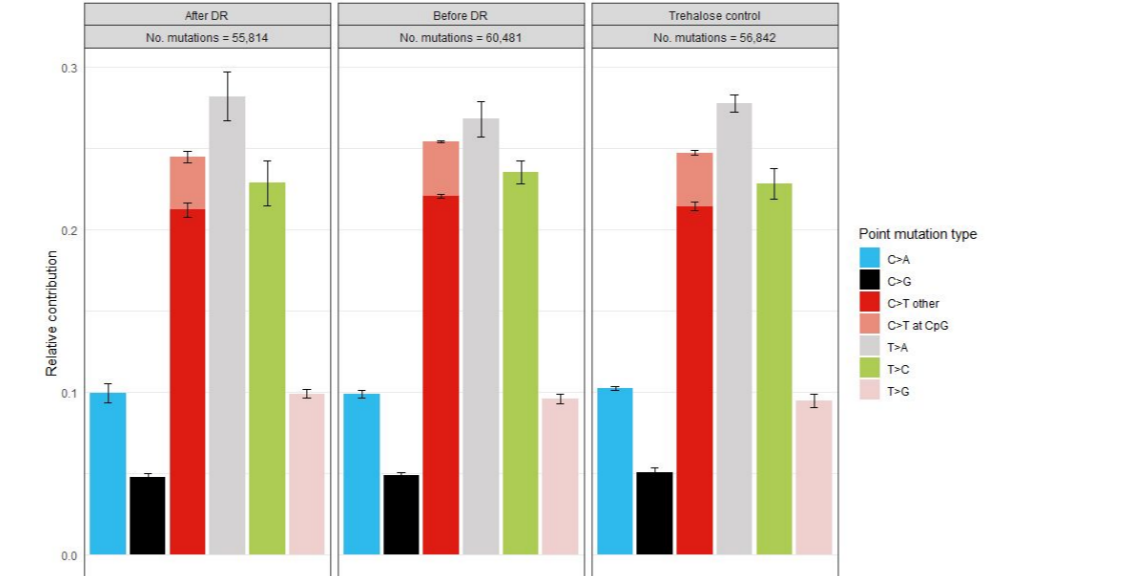


Fig. 2. Mutation spectrum of all experimental group
Рис. 2. Мутационный спектр экспериментальных групп

Table 2. Number of mutation type occurrences in each experiment replica
Таблица 2. Количество замен оснований в каждой экспериментальной реплике

	C>A	C>G	C>T	T>A	T>C	T>G	C>T at CpG	C>T other	Sum
1 Before DR	2333	1117	5982	6613	5486	2192	778	5204	29705
1 After DR	1871	895	4709	4958	4280	1816	614	4095	23238
1 Trehalose control	1781	922	4861	4689	4433	1763	627	4034	22890
2 Before DR	1603	868	4372	4577	4229	1665	630	3742	21686
2 After DR	1954	911	4565	5598	4274	1980	565	4000	23827
2 Trehalose control	1996	859	4687	5580	4220	1896	605	4082	23925
3 Before DR	1607	801	3892	4478	3438	1566	548	3344	19674
3 After DR	2194	1150	5387	5943	4992	1977	693	4694	27030
3 Trehalose control	2013	926	4769	5330	4579	1800	625	4144	24186

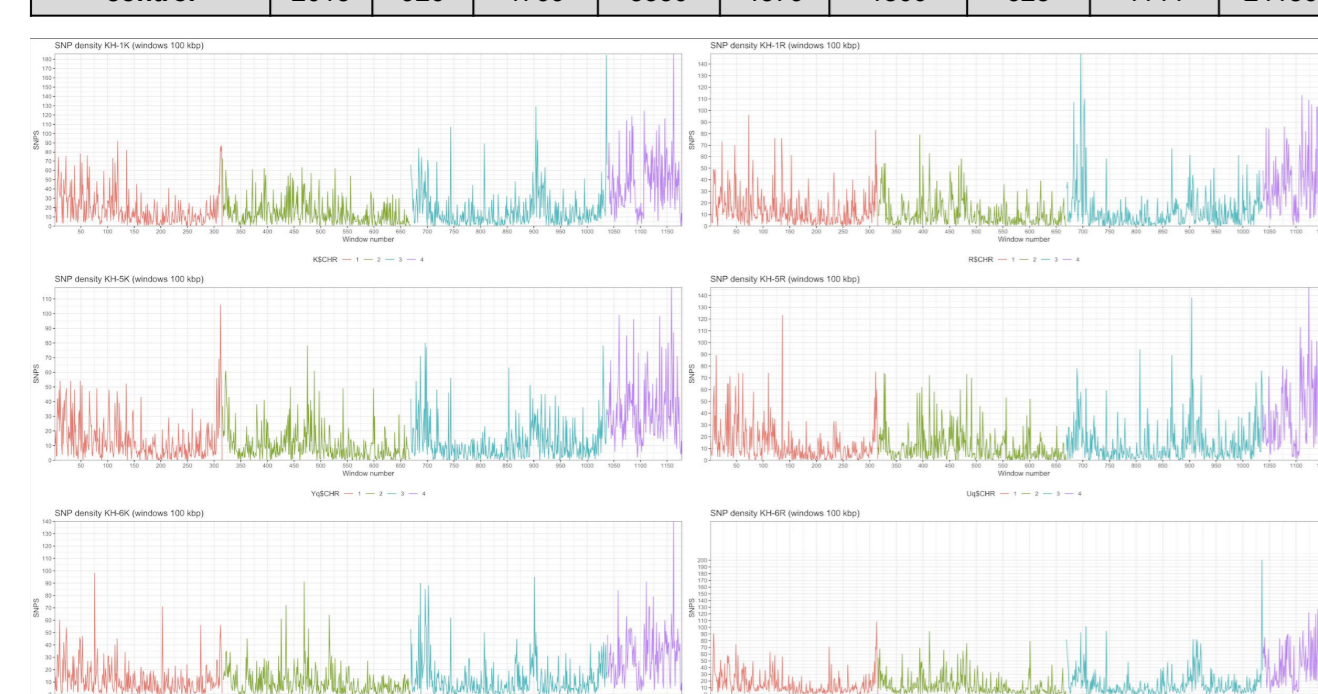


Figure 4. SNPs distribution along genome of each replica
Рис. 4. Распределение SNP вдоль генома каждой реплики

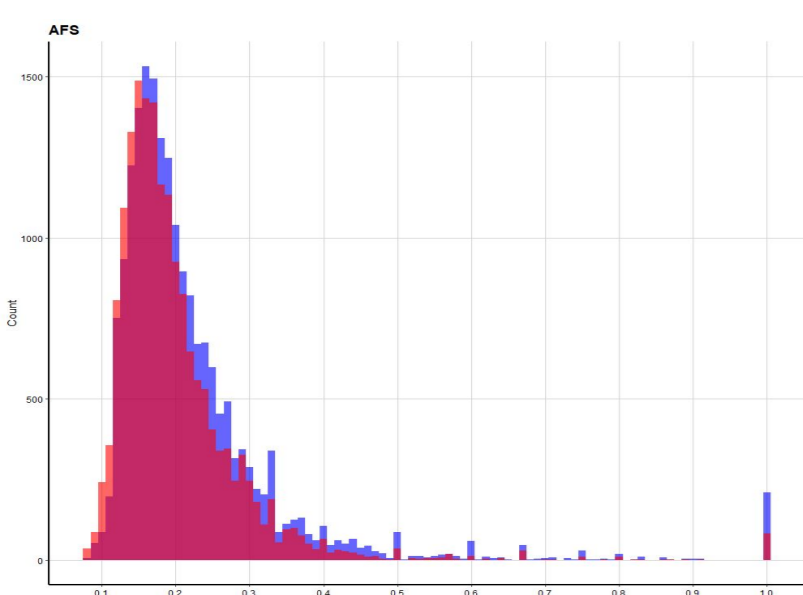


Fig. 6. Allele frequency spectrum
Рис. 6. Спектр частот аллелей

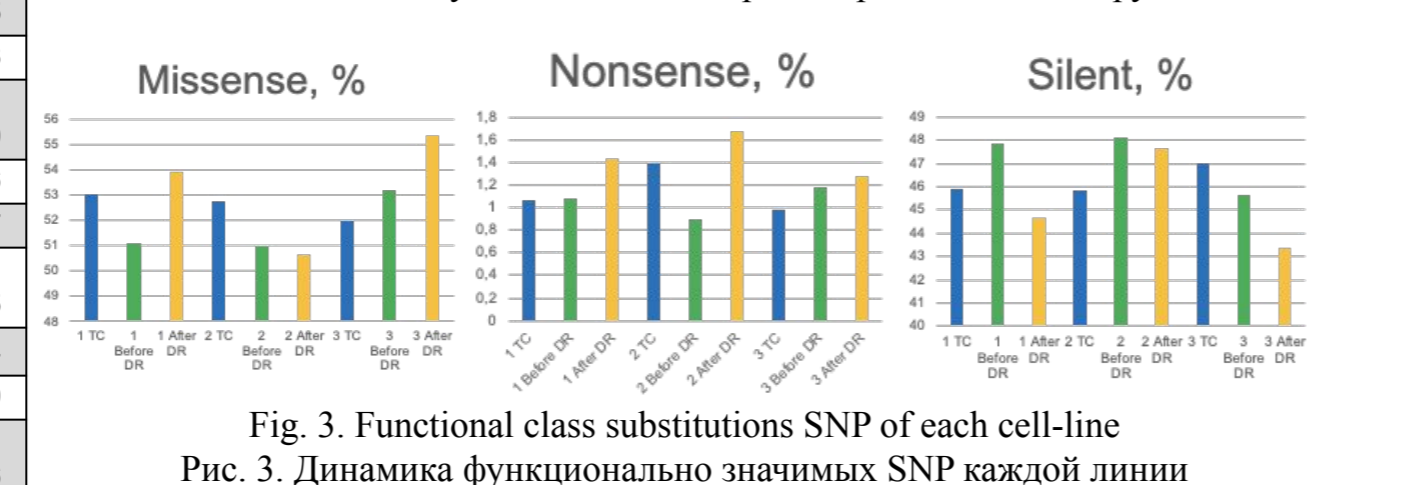


Fig. 3. Functional class substitutions SNP of each cell-line
Рис. 3. Динамика функционально значимых SNP каждой линии

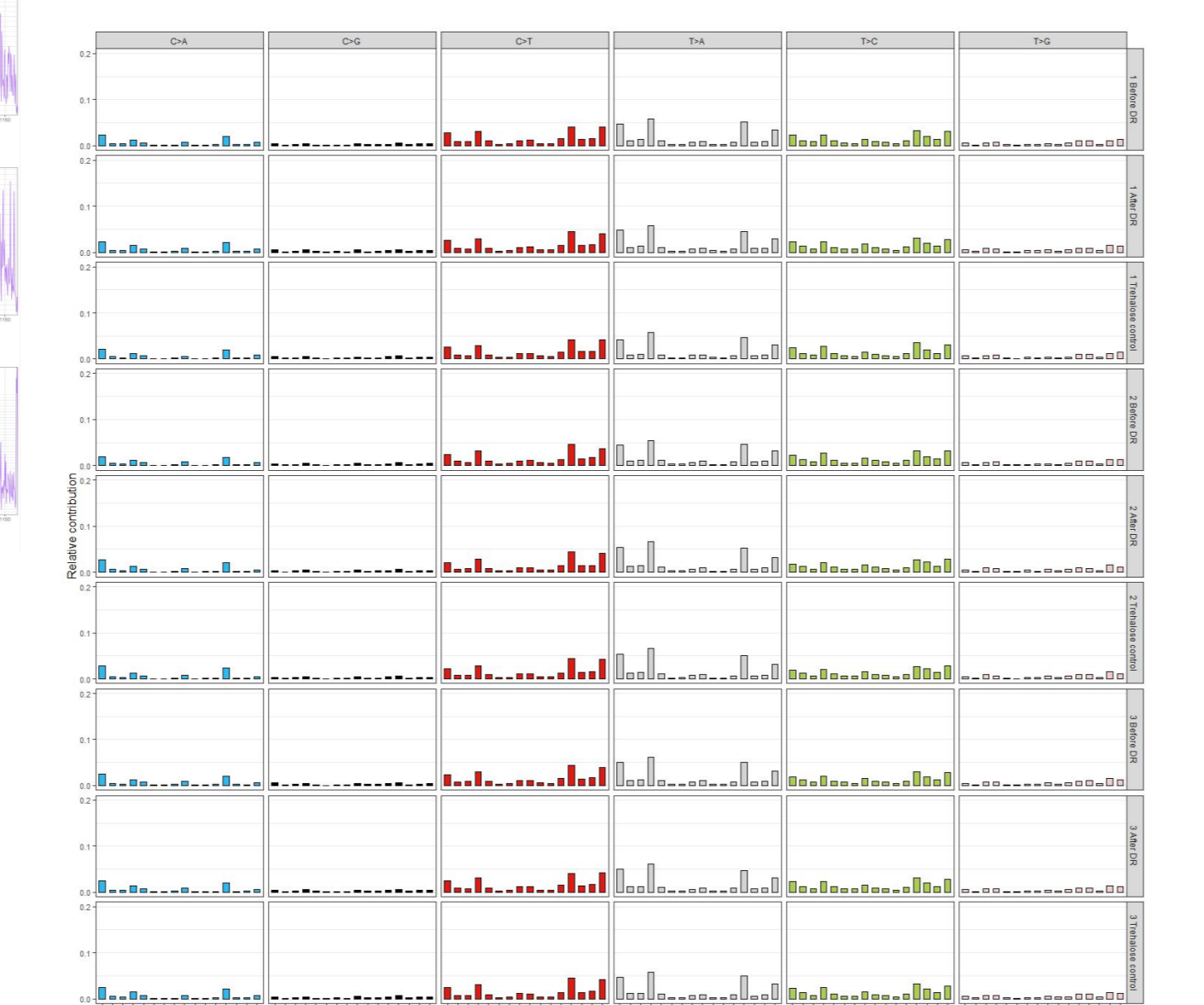


Figure 5. 96-trinucleotide mutational profile
Рис. 5. 96-тринуклеотидный мутационный профиль

2. GENOMIC REARRANGEMENTS AND REPEAT X ARE ASSOCIATED

Large and small genomic rearrangements were analyzed. Common rearrangements were filtered out (table 3). Insertions, deletions, and translocations were found to be more frequent than inversions and duplications (fig. 8). Visual inspection of several rearrangements revealed that the genome contains two types of repeats of different lengths (249 bp and 707 bp) associated with the rearrangements. These repeats were studied in more detail using a variety of algorithms (table 5). No matches for known sequences beyond *P. vanderplanki* were observed. The repeats are associated with at least 4-5% of all rearrangements. While the functional significance of these repeats is not yet clear, most likely they do not encode proteins and are not mobile elements.

Table 3. Total number of SV across all lines
Таблица 3. Общее количество перестроек во всех линиях

	Total number of SVs	Trehalose control	Before DR	After AR
1 Before filtering FP	938	870	1648	
1 After filtering FP	706	597 / 724	1218	
2 Before filtering FP	45	892	3	
2 After filtering FP	31	740 / 734	7	
3 Before filtering FP	952	269	305	
3 After filtering FP	743	214 / 189	232	

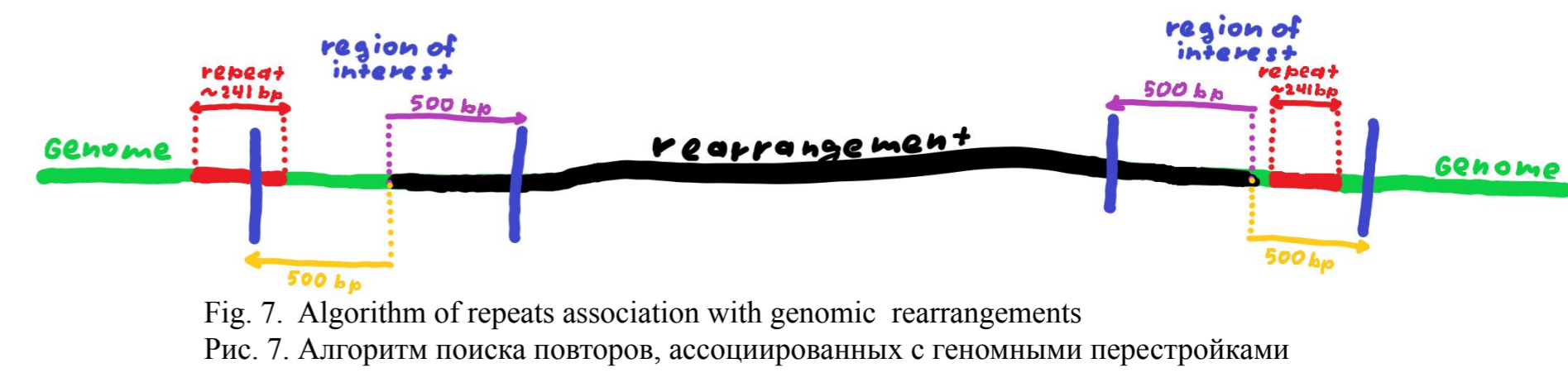


Fig. 7. Algorithm of repeats association with genomic rearrangements
Рис. 7. Алгоритм поиска повторов, ассоциированных с геномными перестройками

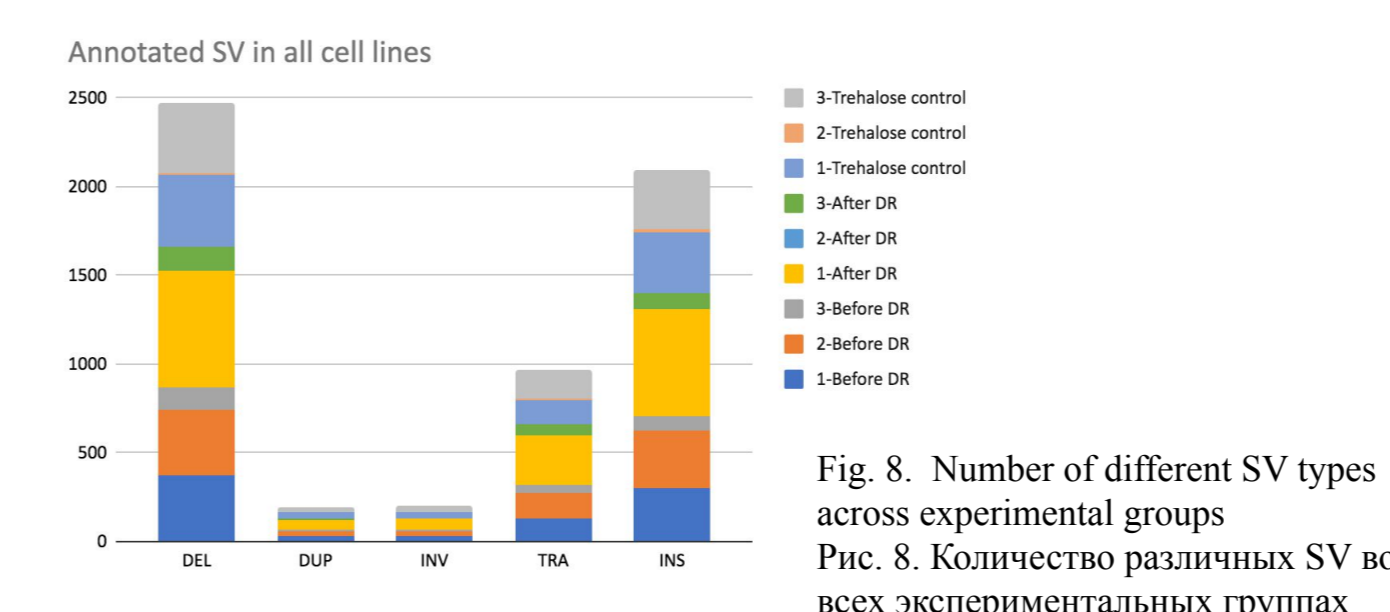


Fig. 8. Number of different SV types across experimental groups
Рис. 8. Количество различных SV во всех экспериментальных группах

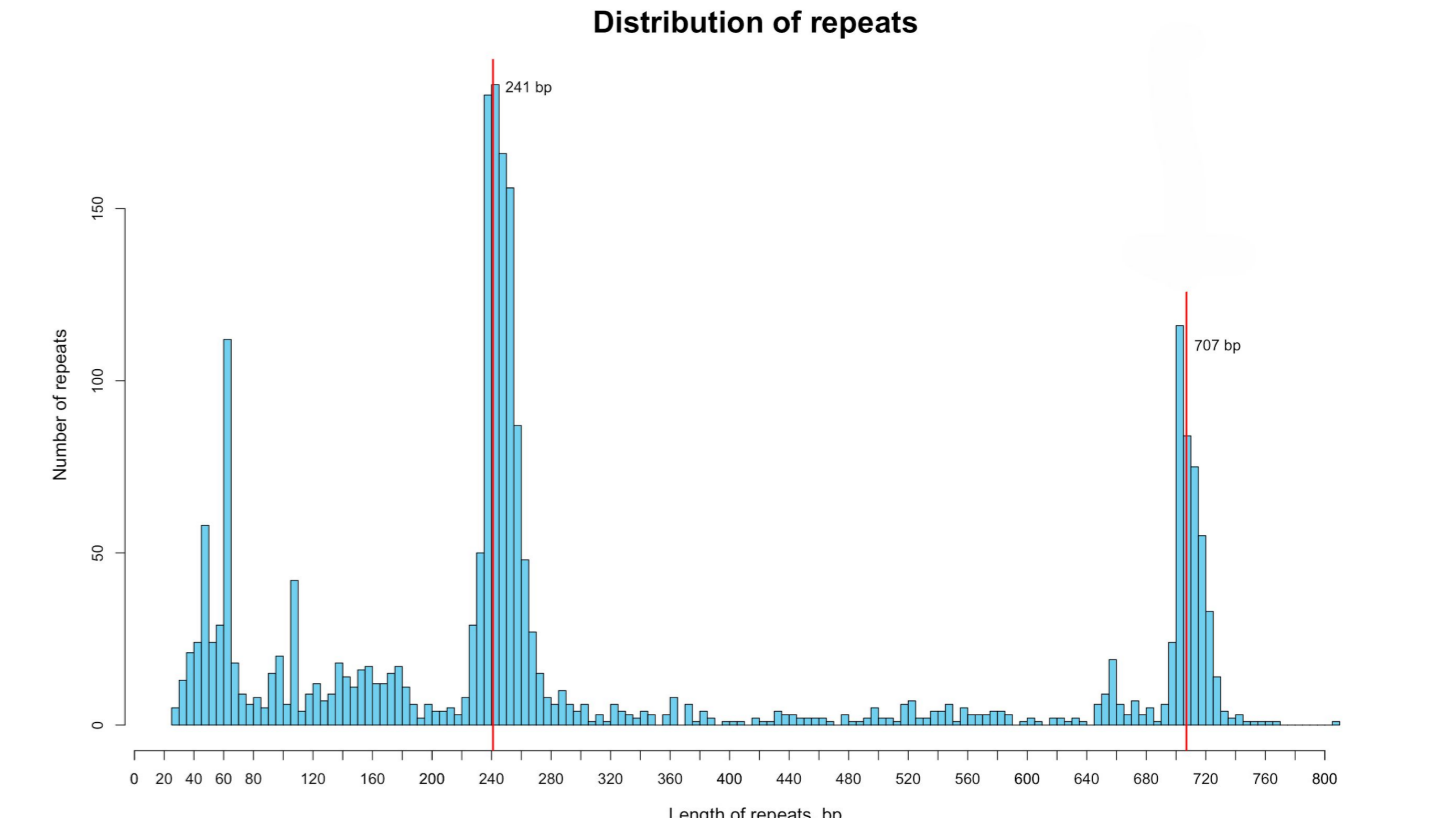


Fig. 9. Distribution of repeats length along midge genome
Рис. 9. Распределение длин повторов по всему геному комара



Fig. 10. RepeatMasker found only simple repeats in our Repeat-X
Рис. 10. RepeatMasker нашел только простые повторы в нашем Повторе-X

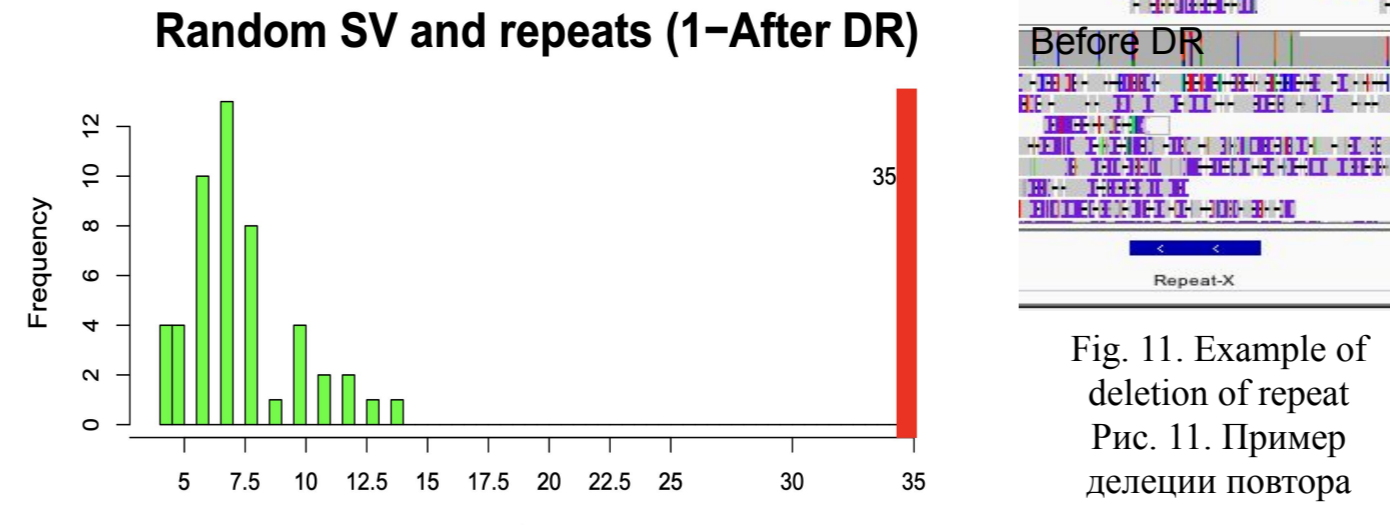
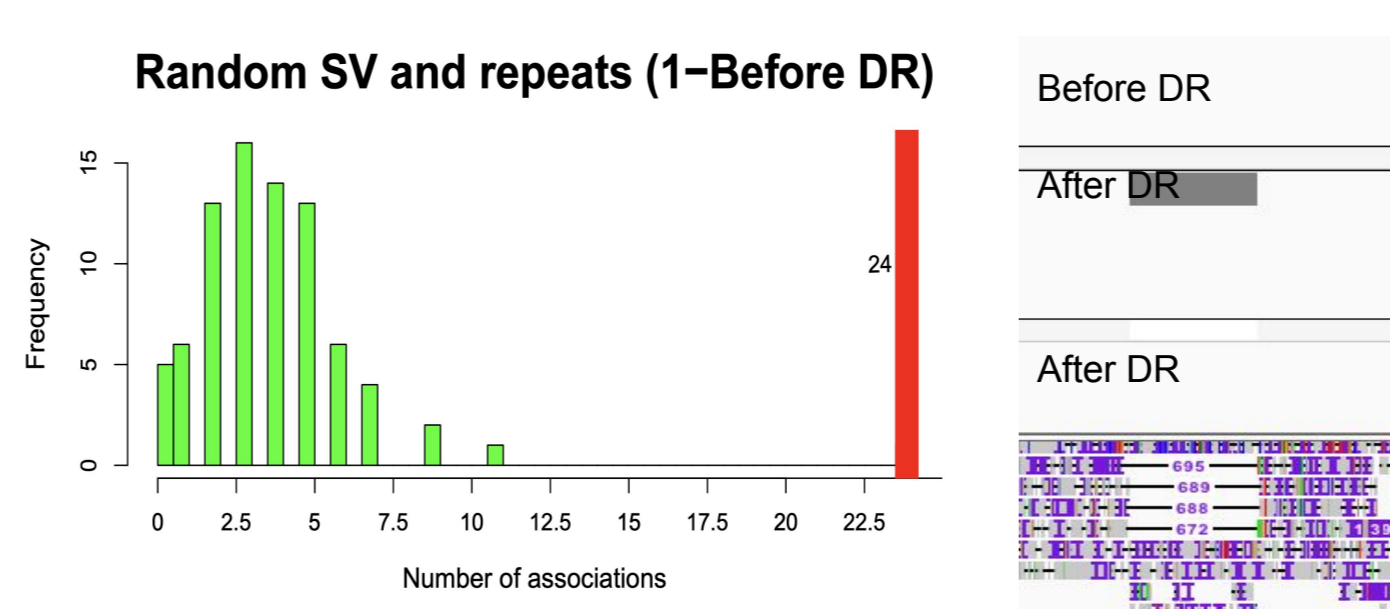


Fig. 11. Example of deletion of repeat
Рис. 11. Пример делеции повтора

Fig. 12. Distribution of random repeats association with random SV (50 iteration) in one replica
Рис. 12. Распределение количества связей случайных повторов и случайных перестроек в одной реплике

Table 4. Distribution of repeats across all chromosomes
Таблица 4. Распределение повторов по хромосомам

	All repeats	Small repeat (250-270)	Big repeat (690-730)	Ratio small
chr3	605	259	93	42.8%
chr2	587	246	111	41.9%
chr1	683	261	134	37.6%
chr4	335	126	64	37.6%
total	2220	892	402	40.1%

Table 5. Methods that were used for finding nature of repeats

Repeats investigation	
Method	Result
Repeat Masker	Simple repeat
Rfam	-
blastx	-
blastn	-
Ensembl Metazoa	-
Transcriptomic analysis	Not done

Table 6. Repeats associated with different types of mutations
Таблица 6. Повторы ассоциированные с разными типами мутаций

	1 Before DR	2 Before DR	3 Before DR	1 After DR	2 After DR	3 After DR	1 Trehalose control	3 Before DR	3 Trehalose control
Deletions	6	9	5	15	0	3	52	0	50
Insertions	16	10	2	19	0	2	48	2	82
Inversions	0	0	0	0	0	0	0	0	2
Duplications	2	2	0	1	1	0	0	1	5
Total number of repeat association with SV	24	21	7	35	1	5	100	2	140

CONCLUSIONS

- Мы не видим значительных изменений в профиле мутации клеточной линии до и после цикла сушки-регидратации, но среднее количество SNP выше в клетках после сушки.
- Мы наблюдаем геномные перестройки, которые произошли после цикла сушки-регидратации в двух репликах, но по сравнению с данными контроля (трегалоза без сушки) нет никакой разницы.
- Идентифицировано два семейства повторов ассоциированных с геномными перестройками.
- Частота SNP значительно выше в регионах вблизи концов перестроек.
- The average number of SNP is higher in cells after the desiccation-rehydration cycle, but there are no significant differences in the mutation profile of the midge cell line before and after the cycle
- We observe genomic rearrangements that happened after desiccation-rehydration cycle in two replicas, but there are no discernible differences with the control data of trehalose treatment without desiccation.
- Two new families of repeats associated with genomic rearrangements have been identified
- The frequency of SNP is considerably higher in regions near rearrangement breakpoints

PERSPECTIVES

- Analyze the evolutionary origin and functional role of repeats and their possible involvement in genomic rearrangements.
- Redo the experiment (to explain the mysteriously high mutation rate in the trehalose control group and ameliorate the low average coverage in second technical replica). We need to increase the time of anhydrobiosis and increase the number of cycles for better data.
- Improve the rearrangement filtration procedure.

ACKNOWLEDGMENTS

We thank Natalia Gogoleva, Yuriy Gogolev, Sabina Kondratieva, Leila Shigapova, Maria Logacheva, Sofya Garushyants, Georgii Bazykin and Oleg Gusev for wet part of this project and their help. This work was made possible by grant from Skolkovo Institute of Science and Technology.