INDELs as genetic markers for local adaptation

Samuel Perini et al.

6th September 2019

# Questions

Can short INDELs be used as additional genetic markers for patterns of local adaptation and ecological speciation? How does the pattern of INDEL variation differ from the one of SNP variation in *L. saxatilis* (compare site frequency spectra and Tajima’s D, outlier positions, correlation of density per window)?  
EXTRA: Are non-synonymous SNPs flanked by frameshift INDELs (epistasis)?

# Hypothesis

Nucleotide divergence between locally adapted populations can be captured using single base variants because these can be reliable genetic markers for both patterns of neutral and adaptive differentiation. SNPs are commonly called together with short INDELs but the latter are often removed because their expected low frequency is not suitable for studying the dynamics of population divergence. However, short INDELs can still be subject to the same population genomic processes as SNPs and, more importantly, being informative for additional evolutionary patterns that have occurred during local adaptation and ecological speciation. For instance, fixed differences between closely related species as well as polymorphisms within species can be composed of a higher percentage of short INDELs than SNPs (Britten 2002; Britten et al. 2003). Therefore, including INDELs as additional genetic markers can be instrumental for population genomic studies because they can represent a significant part of genetic variation.

# Background

1. Mutations are changes in the nucleotide sequence of the DNA and without them evolution would not exist because traits can change over time only if there is genetic variation for genetic drift and natural selection to act upon. Mutations occur in different sizes and as they go from one base to (size of the longest variant), their contribution to genetic diversity increases as well as their potential to decrease fitness of the individuals carrying the mutation. Accordingly, different species and also distinct forms of the same species (ecotypes) have been found to differ for a greater number of single nucleotides than for long structural variants. The explanation behind this observation is almost intuitive, large modifications to a functional system are more likely to cause damage than small ones and thus, they will be maintained at lower frequency (Massouras et al. 2012). However, there are examples of non-SNP variants with positive effects such as … (human unique traits Chen et al. 2007 and others in specialissuesvs).  
   Originally, mutations were detected using (first method of variant calling) and now, after the emergence of high-throughput and next-generation sequencing technologies, it is affordable to scan whole genomes in search for different types and lengths of mutations. Single nucleotide polymorphisms (SNPs) have been primary targets of genetic and genomic analyses that aimed to … (applications for SNPs (Chen et al. 2009; Brumfield et al. 2003; Morin et al. 2004)). Few studies have also expanded their analysis to other types of genetic variants in order to establish a more extensive catalog of genomic variation for the inverstigation of patterns of evolutionary change (specialissuesvs; Chakraborty et al. 2018). In particular, studies on short insertions and deletions (INDELs 50 bp) … (Chen et al. 2009; Barton and Zeng 2019). Therefore, evolutionary geneticists have already acknowledged the role of INDELs in respect to … (Chen et al. 2009) but our knowledge is still limited to a few model species.
2. Studies on humans (Consortium and others 2010) … and other model species (livestock Kang et al. 2015)… were the first ones to assess INDEL variation and they were also the first ones to encounter the challenges associated with INDEL discovery (Väli et al. 2008; Onishi-Seebacher and Korbel 2011). Mapping algorithms deal poorly with long INDELs and with repeated motifs (Narzisi and Schatz 2015). False-negatives and false-positives can be generated when coverage distribution is not uniform or the efficiency of targeted resequencing is not even across the queried regions of the genome (Fang et al. 2016). Currently, the best practice is to … (GATK best practice; Li et al. 2018 and more).
3. Calling INDELs and SNPs have revealed the tendency for these variants to form clusters along the genome and for this reason combining INDELs and SNPs can increase the power to analyse a certain evolutionary pattern (Huang et al. 2014 for a list of studies about SNPs and INDELs clustering). The major histocompatibility complex in birds is one example where an excess of non-synonymous substitutions corresponded to an increase in frequency of deletions (Minias et al. 2018). Other examples involve … (swine Kang et al. 2015; fruit fly Huang et al. 2014). This co-occurrence of SNPs and INDELs along the genome can be a consequence of three different mechanisms (Jovelin and Cutter 2013; Huang et al. 2014). INDELs themselves can be mutagenic (the eukaryote genome contains mutation hotspots where the occurrence of an INDEL may have increased the rate of nucleotide substitution in its sorrounding (Tian et al. 2008); the numerous forms of the major histocompatibility complex in vertebrates might have originated under this mechanism (Minias et al. 2018)), the regions where they fall can be mutagenic and/or the same evolutionary processes operate on both INDELs and SNPs.  
   On the other hand, INDEL variation can carry unique information on how different species or distinct populations of the same species have evolved diverse genomes (Gregory 2004). In fact, INDELs are typically employed in studies of genome evolution because … (Barton and Zeng 2018, 2019; hollister2009; Petrov et al. 2000; Huang et al. 2014 but see @ellis2014). Furthermore, there is evidence for INDELs to have induced the development of human specific traits (Chen et al. 2007) and contributed to phenotypic variance in the fission yeast by interfering at the gene regulatory level (Jeffares et al. 2015). More studies on intraspecific INDEL variation (Chen et al. 2019). In principle, intraspecies variation of INDEL density along the genome may represent a significant source of genetic variation, part of it being adaptive and involved in the process of local adaptation and ecological speciation.
4. SNP-based study and genome scan (Galindo, Grahame, and Butlin 2010; Westram et al. 2018) in *Littorina saxatilis*, species biology and results.

# Materials and methods

Called variants including outgroup as this allows to polarise those mutations that are fixed between species and polymorphic between ecotypes of the same species.

EXTRA: Phasing data using pair-end reads to obtain haplotype sequences and be able to measure haplotype diversity. Mutations are expected to accumulate nearby frameshift INDEL if this is beneficial (shelter load).

# References

Barton, Henry J, and Kai Zeng. 2018. “New Methods for Inferring the Distribution of Fitness Effects for Indels and Snps.” *Molecular Biology and Evolution* 35 (6): 1536–46.

———. 2019. “The Impact of Natural Selection on Short Insertion and Deletion Variation in the Great Tit Genome.” *Genome Biology and Evolution* 11 (6): 1514–24.

Britten, Roy J. 2002. “Divergence Between Samples of Chimpanzee and Human Dna Sequences Is 5%, Counting Indels.” *Proceedings of the National Academy of Sciences* 99 (21): 13633–5.

Britten, Roy J, Lee Rowen, John Williams, and R Andrew Cameron. 2003. “Majority of Divergence Between Closely Related Dna Samples Is Due to Indels.” *Proceedings of the National Academy of Sciences* 100 (8): 4661–5.

Brumfield, Robb T, Peter Beerli, Deborah A Nickerson, and Scott V Edwards. 2003. “The Utility of Single Nucleotide Polymorphisms in Inferences of Population History.” *Trends in Ecology & Evolution* 18 (5): 249–56.

Chakraborty, Mahul, Nicholas W VanKuren, Roy Zhao, Xinwen Zhang, Shannon Kalsow, and JJ Emerson. 2018. “Hidden Genetic Variation Shapes the Structure of Functional Elements in Drosophila.” *Nature Genetics* 50 (1): 20.

Chen, Feng-Chi, Chueng-Jong Chen, Wen-Hsiung Li, and Trees-Juen Chuang. 2007. “Human-Specific Insertions and Deletions Inferred from Mammalian Genome Sequences.” *Genome Research* 17 (1): 16–22.

Chen, Feng, Fengling Lai, Majing Luo, Yu-San Han, Hanhua Cheng, and Rongjia Zhou. 2019. “The Genome-Wide Landscape of Small Insertion and Deletion Mutations in Monopterus Albus.” *Journal of Genetics and Genomics* 46 (2): 75–86.

Chen, Jian-Qun, Ying Wu, Haiwang Yang, Joy Bergelson, Martin Kreitman, and Dacheng Tian. 2009. “Variation in the Ratio of Nucleotide Substitution and Indel Rates Across Genomes in Mammals and Bacteria.” *Molecular Biology and Evolution* 26 (7): 1523–31.

Consortium, 1000 Genomes Project, and others. 2010. “A Map of Human Genome Variation from Population-Scale Sequencing.” *Nature* 467 (7319): 1061.

Ellis, Lisa L, Wen Huang, Andrew M Quinn, Astha Ahuja, Ben Alfrejd, Francisco E Gomez, Carl E Hjelmen, et al. 2014. “Intrapopulation Genome Size Variation in d. Melanogaster Reflects Life History Variation and Plasticity.” *PLoS Genetics* 10 (7): e1004522.

Fang, Han, Ewa A Bergmann, Kanika Arora, Vladimir Vacic, Michael C Zody, Ivan Iossifov, Jason A O’Rawe, et al. 2016. “Indel Variant Analysis of Short-Read Sequencing Data with Scalpel.” *Nature Protocols* 11 (12): 2529.

Galindo, J, JW Grahame, and RK Butlin. 2010. “An Est-Based Genome Scan Using 454 Sequencing in the Marine Snail Littorina Saxatilis.” *Journal of Evolutionary Biology* 23 (9): 2004–16.

Gregory, T Ryan. 2004. “Insertion–Deletion Biases and the Evolution of Genome Size.” *Gene* 324: 15–34.

Huang, Wen, Andreas Massouras, Yutaka Inoue, Jason Peiffer, Miquel Ràmia, Aaron M Tarone, Lavanya Turlapati, et al. 2014. “Natural Variation in Genome Architecture Among 205 Drosophila Melanogaster Genetic Reference Panel Lines.” *Genome Research* 24 (7): 1193–1208.

Jeffares, Daniel C, Charalampos Rallis, Adrien Rieux, Doug Speed, Martin Převorovskỳ, Tobias Mourier, Francesc X Marsellach, et al. 2015. “The Genomic and Phenotypic Diversity of Schizosaccharomyces Pombe.” *Nature Genetics* 47 (3): 235.

Jovelin, Richard, and Asher D Cutter. 2013. “Fine-Scale Signatures of Molecular Evolution Reconcile Models of Indel-Associated Mutation.” *Genome Biology and Evolution* 5 (5): 978–86.

Kang, Huimin, Haifei Wang, Ziyao Fan, Pengju Zhao, Amjad Khan, Zongjun Yin, Jiafu Wang, et al. 2015. “Resequencing Diverse Chinese Indigenous Breeds to Enrich the Map of Genomic Variations in Swine.” *Genomics* 106 (5): 286–94.

Li, Donghe, Wonji Kim, Longfei Wang, Kyong-Ah Yoon, Boyoung Park, Charny Park, Sun-Young Kong, et al. 2018. “Comparison of Indel Calling Tools with Simulation Data and Real Short-Read Data.” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*.

Massouras, Andreas, Sebastian M Waszak, Monica Albarca-Aguilera, Korneel Hens, Wiebke Holcombe, Julien F Ayroles, Emmanouil T Dermitzakis, et al. 2012. “Genomic Variation and Its Impact on Gene Expression in Drosophila Melanogaster.” *PLoS Genetics* 8 (11): e1003055.

Minias, Piotr, Ewa Pikus, Linda A Whittingham, and Peter O Dunn. 2018. “A Global Analysis of Selection at the Avian Mhc.” *Evolution* 72 (6): 1278–93.

Morin, Phillip A, Gordon Luikart, Robert K Wayne, and others. 2004. “SNPs in Ecology, Evolution and Conservation.” *Trends in Ecology & Evolution* 19 (4): 208–16.

Narzisi, Giuseppe, and Michael C Schatz. 2015. “The Challenge of Small-Scale Repeats for Indel Discovery.” *Frontiers in Bioengineering and Biotechnology* 3: 8.

Onishi-Seebacher, Megumi, and Jan O Korbel. 2011. “Challenges in Studying Genomic Structural Variant Formation Mechanisms: The Short-Read Dilemma and Beyond.” *Bioessays* 33 (11): 840–50.

Petrov, Dmitri A, Todd A Sangster, J Spencer Johnston, Daniel L Hartl, and Kerry L Shaw. 2000. “Evidence for Dna Loss as a Determinant of Genome Size.” *Science* 287 (5455): 1060–2.

Tian, Dacheng, Qiang Wang, Pengfei Zhang, Hitoshi Araki, Sihai Yang, Martin Kreitman, Thomas Nagylaki, Richard Hudson, Joy Bergelson, and Jian-Qun Chen. 2008. “Single-Nucleotide Mutation Rate Increases Close to Insertions/Deletions in Eukaryotes.” *Nature* 455 (7209): 105.

Väli, Ülo, Mikael Brandström, Malin Johansson, and Hans Ellegren. 2008. “Insertion-Deletion Polymorphisms (Indels) as Genetic Markers in Natural Populations.” *BMC Genetics* 9 (1): 8.

Westram, Anja M, Marina Rafajlović, Pragya Chaube, Rui Faria, Tomas Larsson, Marina Panova, Mark Ravinet, et al. 2018. “Clines on the Seashore: The Genomic Architecture Underlying Rapid Divergence in the Face of Gene Flow.” *Evolution Letters* 2 (4): 297–309.