

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.

For more information visit www.intechopen.com



Landscape Genetics: From Classic Molecular Markers to Genomics

Enéas Ricardo Konzen and Maria Imaculada Zucchi

Abstract

Landscape genetics combines population genetics and landscape ecology to understand processes that shape the distribution and organization of human, animal, or plant populations. This field of genetics emerged from the availability of several studies with classical molecular markers, such as isozymes, RAPD, AFLP, and microsatellites. Population genetic studies enabled the detection of population structure with those markers, but a more comprehensive analysis of natural populations was only possible with the development of statistical methods that combined both molecular data and environmental variables. Ultimately, the rapid development of sequencing technologies allowed studies at the genomic level, augmenting the resolution of association with environment factors. This chapter outlines basic concepts in landscape genetics, the main statistical methods used so far, and the perspectives of this field of knowledge into strategies for conservation of natural populations of plant and animal species. Moreover, we briefly describe the application of the field to understand historical human migration processes as well as how some diseases are spread throughout the world.

Keywords: molecular studies, environmental variables, population structure, genetic diversity, single nucleotide polymorphisms

1. Introduction

Population genetic studies deal with allele frequencies and processes that shape their variation within and among populations. Multiple studies have addressed genetic variation and their structure based on the screening of molecular markers such as allozymes (began with Lewontin and Hubby [1]), random amplified polymorphic DNA (RAPD) [2], amplified fragment length polymorphism (AFLP) [3], microsatellites or simple sequence repeats (SSR) [4], intersimple sequence repeats (ISSR) [5] and single nucleotide polymorphisms (SNP). The use of allozyme markers started up a series of population genetic studies, allowing relatively precise estimation of heterozygosity levels due to their codominance nature. Those markers were largely employed until the end of the 1990s. The development of techniques for screening directly at the level of DNA has accelerated the discovery of numberless markers in humans, animals, plants, fungi, and other organisms. RAPD, ISSR, and AFLP, in general, are more limited in describing genetic variation due to their dominance. In contrast, several SSR markers have been developed for studying a diverse set of species, enabling precise estimates of genetic diversity, gene flow, spatial genetic structure, paternity, linkage, and association mapping.

Ultimately, SNP markers have arisen as powerful markers for fine-scale genetic diversity, structure, and association mapping studies. The direct comparison among sequences of specific fragments generated by Sanger sequencing allowed the discovery of the first set of SNP. However, the revolution in sequencing technology of the last decade has provided numberless sequences for comparing individuals and deciphering population genetic mechanisms with high accuracy. The next-generation sequencing platforms generate millions of sequences that often result in thousands of SNP markers.

Nonetheless, the sole use of molecular data provides no definitive responses on evolutionary mechanisms operating in populations. An examination of the ecological factors, that drive the fate of individuals over generations or how current mechanisms impact in their adaptation or acclimation, is a much-needed task to better understand all species. Adequate statistical methods combining genetic and environmental variables are then necessary. Landscape genetics emerged as a field for the improvement of our understanding of the influence of geographical and environmental variables on the genetic structure of populations [6]. It diverges from the traditional basis of population genetics in the sense of more profound tests of the influence of landscape and environmental factors such as altitude, topography, and ground cover on population processes such as gene flow and population structure [7]. The rapid boost in genome-scale analyses also generated the terminology landscape genomics, as proposed by Joost et al. [8]. Landscape genomics differs from landscape genetics in the sense that it has become a powerful approach for scanning genes involved in complex adaptation mechanisms of species at populations and individual levels [9, 10].

This chapter is intended to provide brief concepts that cover the subject of landscape genetics and genomics. Furthermore, we outline potential applications of landscape genetic studies in the comprehension of adaptive traits of plants and animals and how such results may assist in the design of conservation strategies for endangered species. It is not our intent to provide an exhaustive panorama of landscape genetics studies so far, but rather contextualize concepts and applications with chosen case studies. Moreover, we briefly contextualize how landscape genetics is contributing in the comprehension of historical human migrations and the dispersion of human diseases.

2. Molecular markers and population structure studies

The most popular molecular markers employed in population genetic studies are SSR [4] and SNP. Simple sequence repeats are tandem repeated motifs with 1–6 bp [11] or up to 10 bp [12] with high frequency in genomes of all organisms. Plants commonly have AT-type repeats, whereas animals have the AC motif as the most common repeat unit [13]. High mutation rates are characteristics of microsatellite markers [12] providing markers with several alleles. SSR are codominant, hypervariable, and Mendelian inherited [14], which is implicated in high heterozygosity levels, increasing the discriminatory power among individuals and populations. Originally, SSR were developed from DNA libraries that required extensive laboratory work. Currently, however, the easiest way of discovering novel microsatellites is through direct sequencing of genomes and transcriptomes generated from NGS platforms [12]. With that available, SNP markers have actually been the most studied markers in recent years. SNP markers are the most abundant polymorphisms along plant and animal genomes. SNP consist on single base-pair changes present in the genome sequence that can occur as transitions or transversions, as nucleotide substitutions [15]. They can reach much higher density than all other types of markers in

genomes. Next-generation sequencing can generate large amounts of sequence data, enabling the detection of thousands of SNP [16].

Microsatellites and SNP markers are powerful tools for population genetic analyses. They have been extensively employed in studies with humans as well as animal and plant models and non-model species. The codominance and multiallelic nature of microsatellites make them suitable for estimating variables such as heterozygosity, inbreeding, gene flow, outcrossing rates, differentiation among populations and population structure [17]. SNP markers are generally employed for determining population structure as well, but with much higher density of markers and therefore genomic coverage to explain such subdivision. A series of studies have used SNP to dissect complex traits with QTL mapping and genome-wide association studies (GWAS) [15].

3. The concepts of landscape genetics and landscape genomics

Landscape genetics is concerned with testing the effects of landscape features on gene flow and genetic population structure. In general, the first studies of landscape genetics involved an exploratory phase, by geographically widespread sampling of populations and analysis of the effects of various landscape variables [18]. Landscape features or variables consist of any biotic, climatic, soil, or other conditions that comprise the habitat of organisms [6]. The population structure means the organization of genetic variation as influenced by a combination of evolutionary forces such as recombination, mutation, drift, natural selection, and historic demographic processes [19]. This leads to the idea that a group of subpopulations that exchange migrants in an occasional fashion are part of metapopulations [6].

The current status of genomic technologies allows the discovery of thousands of SNP markers, which has increased the resolution power for studying the association of environmental variables with specific genomics regions, also with a much deeper understanding of evolutionary processes. Genotyping-by-sequencing has enabled the discovery of SNP markers even in non-model species, which may lack a reference genome so far [20, 21]. This is where the concept of landscape genomics comes forward. Landscape genomics focuses on detecting candidate genes under selection as putative signals of local adaptation. The design of a landscape genomics experiment involves replicated sampling of environmental factors that might be driving selection, augmenting the resolution for detection of candidate loci under selection [10].

4. A briefing on statistical approaches in landscape genetics

In a landscape genetics study, two steps of analyses are normally required. The first involves the analysis of patterns of genetic variation. Next, such patterns are correlated with landscape variables based on statistical methods [22]. To test for association of environmental variables with genetic data, one of the simplest and commonly used methods is the Mantel's test, originally developed for identifying time-space clustering of diseases [23]. The test uses permutations to address the significance of the linear correlation coefficient between two pair-wise similarity or dissimilarity matrices [22]. One of the simplest examples of its application in landscape genetics is to correlate the genetic distances between individuals with their geographic location [24].

The methods for determining association of genetic data with environmental variables can be broadly categorized into approaches that deal with (i) pair-wise

landscape data and (ii) location-specific landscape data, as reviewed by Balkenhol et al. [22]. The development of methods in landscape genomics, however, expanded the range of tests for detecting loci under selection using genome scans, approaches for candidate gene discovery, QTL mapping and GWAS. Genome scans use two methods for detecting loci under selection, the differentiation outlier methods and the genetic-environmental association test, as reviewed by Storfer et al. [10]. Novel methods are continuously being developed, as more genomes are becoming sequenced or resequenced in populations.

5. Applications of landscape genetics

Several applications of landscape genetics or genomics can be described. We briefly account for case studies in plant and animal systems within this section. Moreover, a few examples of studies applied to humans are also given. In general, landscape genetics or genomics studies have provided association among geographic, abiotic, and biotic factors and genetic data provided by the screening of molecular markers in populations of diverse organisms. It has increased our power to detail inferences of movement and gene flow and potential adaptation to the landscape populations occur. However, studies for several organisms are still scarce or inexistent.

Cultivated crops such as maize, soybean, rice, and common bean were domesticated from wild progenitors which reflect their current adaptation to distinct environments. Landscape genomics studies have enabled a deeper understanding of processes shaping their distribution across multiple environments. Common bean (*Phaseolus vulgaris* L.) is an exceptional example of a widespread species original from America. Molecular data of wild germplasm identified two major gene pools, the Andean from Argentina to Colombia, and the Mesoamerican from Colombia to Mexico [25, 26]. A third smaller pool of wilds is also distinctive in a narrow area between Peru-Ecuador [27]. Microsatellites markers were broadly used to screen the genetic structure of wild and domesticated accessions of common bean (*Phaseolus vulgaris* L.), distinguishing from the broadest Andean and Mesoamerican gene pools to further subdivision within each one of them [25]. SNP markers from single fragments sequenced by Sanger also allowed an accurate distinction between Andean and Mesoamerican accessions, as well as their subdivisions [28]. The recognition of a parallel domestication event in each of the two major pools was also possible based on the detection of SNP markers in specific genomic regions of Andean and Mesoamerican genotypes [29]. Recent landscape genomics approaches enabled a more detailed description of the major events that determined the range expansion of *P. vulgaris* in America and how they were accompanied by environmental changes [26]. The climatic variability was also associated with differential drought adaptation and specific SNP markers were statistically related to root and shoot traits varying in a Mesoamerican panel of genotypes originated from regions with distinct precipitation regimes throughout the year [30].

Another application of landscape genomics concerns with the understanding of range expansion and ecological dominance of insect pests. The first step toward that is to know the population structure, gene flow and how natural selection is affecting adaptation. Zucchi et al. [31] described and addressed such problem by examining the population structure of *Piezodorus guildinii*, a soybean pest, in the United States and Brazil. A GBS-based set of SNP markers revealed genetic structure according to their geographic environment of origin. About 10% of loci were under positive selection, and their annotation revealed genes involved in genome

reorganization, neuropeptides, and energy mobilization [31]. Addressing such problem is to assist future endeavors at managing pest spreading in cultivated crops.

Another equally important questions addressed by landscape genomics are the consequences of climate change and human intervention to natural populations of wilds plants and animals. *Euterpe edulis* Martius is a palm species native to the Atlantic Rain Forest in Brazil, known as heart-of-palm [32]. The species is the list of endangered species to extinction [33]. Several studies have addressed the genetic diversity and structure of natural populations of this palm (for a compilation see [34]). Soares et al. [35] studied the genetic diversity and structure of remnant fragments of *E. edulis* in Bahia state and related the data to landscape metrics such as composition and configuration and local variables including the logging activity as human disturbance variable. No evidence of spatial genetic structure was detected, but distinct genetic clusters could be identified, suggesting a reduction in gene flow between the fragments of this study [35]. Natural populations located in other regions of Brazil, such as in Sao Paulo state, revealed to have high genetic diversity, as shown from microsatellite markers. Adjacent populations that have been generated though germplasm collection for management and cultivation showed similar genetic diversity. Those genetic materials could be used for recovering overexploited populations [36].

Landscape genetics studies with wild animals have been focused in recognizing their patterns of moving across their habitats. On terrestrial lands, landscape genetics of animals has particular features in comparison to aquatic environments or even to terrestrial plants. Landscape patterns interfere with organism behavior, thereby affecting mating and dispersal and reflecting on population processes [37].

6. Landscape genetics and human populations and diseases

Genomic technologies have also enabled studies to uncover historical human migrations and the genetic structure and diversity of human populations. For example, a genome-wide study of Malaysian ethnic groups using a SNP array revealed that humans from the peninsular area of Malaysia had higher genetic diversity, which the authors associated with a contact zone for recent human migrations in the Asian continent [38]. Such an example suggests the association between the genetic structure of human populations with geographic variables. In fact, Peter et al. [39] show that genetic differentiation generally tends to increase over higher geographic distances; however, distortions in those patterns also frequently occur. The human population structure, then, seems to be quite dynamic.

Landscape genetics also has been employed in epidemiological studies of human diseases. Statistical methods can be used in the identification of hotspot areas of disease movement [40]. This will have important implications in designing strategies for spread containment. One challenge, however, has been the application of landscape genetics methods in vector-borne diseases, which was reviewed by Hemming-Schroeder [40]. A few studies have been dedicated to such goal with human diseases. One interesting example is the correlation found between the genetic structure of *Aedes mcintochi*, a major vector for Rift Valley fever in Kenya, and mean precipitation values [41].

In 2020, one of the major global health issues concerns the new COVID-19. Sequencing technologies coupled with landscape genomics approaches have the potential to identify dispersal patterns of the virus in order to contain its spreading. Landscape genetic approaches have the power of assisting the decision-making process.

7. Concluding remarks

Climate change and human interference are no longer to be neglected on natural ecosystems. Among several fields of study devoted to deciphering the impact of these processes, landscape genetics will provide a better comprehension of the interaction between organisms and their environment of origin. The boost in sequencing technologies is enabling the study of the most diverse range of organisms. In fact, the Earth BioGenome Project is intended to sequence, catalog, and characterize all eukaryotic diversity in the forthcoming decade [42]. With that information available, resequencing to the level of population and their association with landscape variables will provide information for designing appropriate strategies for the conservation of endangered forms of life as well as any other species. The resequencing of several human genomes will also enable a better comprehension of the human population structure throughout the world and how the landscape shapes its organization. This has been and will be continuing valuable information to comprehending the dispersion of human diseases as well.

Conflict of interest

The authors declare no conflict of interest.

Author details

Enéas Ricardo Konzen^{1*} and Maria Imaculada Zucchi^{2,3}

1 Interdisciplinary Department, Center of Limnological, Coastal and Marine Studies, Federal University of Rio Grande do Sul, Imbé, Rio Grande do Sul, Brazil

2 Secretary of Agriculture and Supply of São Paulo State, Piracicaba, São Paulo, Brazil

3 Institute of Biology, University of Campinas, Campinas, São Paulo, Brazil

*Address all correspondence to: erkonzen@gmail.com

IntechOpen

© 2020 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Lewontin RC, Hubby JL. A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics*. 1966;**54**(2):595
- [2] Williams JG, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*. 1990;**18**(22):6531-6535
- [3] Vos P, Hogers R, Bleeker M, Reijmans M, Lee TV, Hornes M, et al. AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Research*. 1995;**23**:4407-4414
- [4] Litt M, Luty JA. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human Genetics*. 1989;**44**(3):397
- [5] Lagercrantz U, Ellegren H, Kakanuga T. The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. *Nucleic Acids Research*. 1993;**21**:1111-1115
- [6] Manel S, Schwartz MK, Luikart G, Taberlet P. Landscape genetics: Combining landscape ecology and population genetics. *Trends in Ecology & Evolution*. 2003;**18**(4):189-197
- [7] Storfer A, Murphy MA, Evans JS, Goldberg CS, Robinson S, Spear SF, et al. Putting the 'landscape' in landscape genetics. *Heredity*. 2007;**98**(3):128
- [8] Joost S, Bonin A, Bruford MW, Després L, Conord C, Erhardt G, et al. A spatial analysis method (SAM) to detect candidate loci for selection: Towards a landscape genomics approach to adaptation. *Molecular Ecology*. 2007;**16**(18):3955-3969
- [9] Li Y, Zhang XX, Mao RL, Yang J, Miao CY, Li Z, et al. Ten years of landscape genomics: Challenges and opportunities. *Frontiers in Plant Science*. 2017;**8**:2136
- [10] Storfer A, Patton A, Fraik AK. Navigating the interface between landscape genetics and landscape genomics. *Frontiers in Genetics*. 2018;**9**:68
- [11] Kalia RK, Rai MK, Kalia S, Singh R, Dhawan AK. Microsatellite markers: An overview of the recent progress in plants. *Euphytica*. 2011;**177**(3):309-334
- [12] Vieira MLC, Santini L, Diniz AL, Munhoz CDF. Microsatellite markers: What they mean and why they are so useful. *Genetics and Molecular Biology*. 2016;**39**(3):312-328
- [13] Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, et al. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding*. 1996;**2**(3):225-238
- [14] Abdul-Muneer PM. Application of microsatellite markers in conservation genetics and fisheries management: Recent advances in population structure analysis and conservation strategies. *Genetics Research International*. 2014;**2014**:1-11
- [15] Nadeem MA, Nawaz MA, Shahid MQ, Doğan Y, Comertpay G, Yıldız M, et al. DNA molecular markers in plant breeding: Current status and recent advancements in genomic selection and genome editing. *Biotechnology & Biotechnological Equipment*. 2018;**32**(2):261-285

- [16] Xia W, Luo T, Zhang W, Mason AS, Huang D, Huang X, et al. Development of high-density SNP markers and their application in evaluating genetic diversity and population structure in *Elaeis guineensis*. *Frontiers in Plant Science*. 2019;**10**:130
- [17] Konzen ER. Towards conservation strategies for forest tree endangered species: The meaning of population genetic statistics. *Advances in Forestry Science*. 2014;**1**(1):45-51
- [18] Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP. Landscape genetics: Where are we now? *Molecular Ecology*. 2010;**19**(17):3496-3514
- [19] Andam CP, Challagundla L, Azarian T, Hanage WP, Robinson DA. Population structure of pathogenic bacteria. In: *Genetics and Evolution of Infectious Diseases*. Amsterdam: Elsevier; 2017
- [20] Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*. 2011;**6**(5):e19379
- [21] Poland JA, Brown PJ, Sorrells ME, Jannink JL. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS One*. 2012;**7**(2):e32253
- [22] Balkenhol N, Gugerli F, Cushman SA, Waits LP, Coulon A, Arntzen JW, et al. Identifying future research needs in landscape genetics: Where to from here? *Landscape Ecology*. 2009;**24**(4):455
- [23] Mantel N. The detection of disease clustering and a generalized regression approach. *Cancer Research*. 1967;**27** (2 part 1):209-220
- [24] Diniz-Filho JAF, Soares TN, Lima JS, Dobrovolski R, Landeiro VL, Telles MPDC, et al. Mantel test in population genetics. *Genetics and Molecular Biology*. 2013;**36**(4):475-485
- [25] Kwak M, Gepts P. Structure of genetic diversity in the two major gene pools of common bean (*Phaseolus vulgaris* L., Fabaceae). *Theoretical and Applied Genetics*. 2009;**118**(5):979-992
- [26] Ariani A, Mier B, Teran JC, Gepts P. Spatial and temporal scales of range expansion in wild *Phaseolus vulgaris*. *Molecular Biology and Evolution*. 2017;**35**(1):119-131
- [27] Gepts P, Osborn TC, Rashka K, Bliss FA. Phaseolin-protein variability in wild forms and landraces of the common bean (*Phaseolus vulgaris*): Evidence for multiple centers of domestication. *Economic Botany*. 1986;**40**(4):451-468
- [28] Bitocchi E, Nanni L, Bellucci E, Rossi M, Giardini A, Zeuli PS, et al. Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. *Proceedings of the National Academy of Sciences*. 2012;**109**(14):E788-E796
- [29] Bitocchi E, Bellucci E, Giardini A, Rau D, Rodriguez M, Biagetti E, et al. Molecular analysis of the parallel domestication of the common bean (*Phaseolus vulgaris*) in Mesoamerica and the Andes. *New Phytologist*. 2013;**197**(1):300-313
- [30] Teran JCBM, Konzen ER, Medina V, Palkovic A, Ariani A, Tsai SM, et al. Root and shoot variation in relation to potential intermittent drought adaptation of Mesoamerican wild common bean (*Phaseolus vulgaris* L.). *Annals of Botany*. 2019;**124**(6):917
- [31] Zucchi MI, Cordeiro EM, Allen C, Novello M, Viana JPG, Brown PJ, et al. Patterns of genome-wide variation, population differentiation and SNP discovery of the red banded stink

- bug (*Piezodorus guildinii*). *Scientific Reports*. 2019;**9**(1):1-11
- [32] Reis MS, Guerra MP, Nodari RO, Ribeiro RJ, Reis A. Distribuição geográfica e situação atual das populações na área de ocorrência de *Euterpe edulis* Martius. *Sellowia*. 2000a;**49-52**:324-335
- [33] Martinelli M, Moraes A. Livro Vermelho da Flora Do Brasil. Rio de Janeiro: Instituto de Pesquisas Jardim Botânico do Rio de Janeiro; 2013. p. 1100
- [34] Konzen ER, Martins MP. Contrasting levels of genetic diversity among populations of the endangered tropical palm *Euterpe edulis* Martius. *Cernea*. 2017;**23**(1):31-42
- [35] Soares LASS, Cazetta E, Santos LR, França DDS, Gaiotto FA. Anthropogenic disturbances eroding the genetic diversity of a threatened palm tree: A multi-scale approach. *Frontiers in Genetics*. 2019;**10**:1090
- [36] Novello M, Viana JPG, Alves-Pereira A, de Aguiar Silvestre E, Nunes HF, Pinheiro JB, et al. Genetic conservation of a threatened Neotropical palm through community-management of fruits in agroforests and second-growth forests. *Forest Ecology and Management*. 2018;**407**:200-209
- [37] Waits LP, Cushman SA, Spear SF. Applications of landscape genetics to connectivity research in terrestrial animals [chapter 12]. In: Balkenhol N, Cushman SA, Storfer AT, Waits LP, editors. *Landscape Genetics: Concepts, Methods, Applications*. 1st ed. West Sussex: John Wiley and Sons Ltd; 2016. pp. 199-219
- [38] Deng L, Hoh BP, Lu D, Fu R, Phipps ME, Li S, et al. The population genomic landscape of human genetic structure, admixture history and local adaptation in peninsular Malaysia. *Human Genetics*. 2014;**133**(9):1169-1185
- [39] Peter BM, Petkova D, Novembre J. Genetic landscapes reveal how human genetic diversity aligns with geography. *Molecular Biology and Evolution*. 2019;**37**(4):943-951
- [40] Hemming-Schroeder E, Lo E, Salazar C, Puente S, Yan G. Landscape genetics: A toolbox for studying vector-borne diseases. *Frontiers in Ecology and Evolution*. 2018;**6**:21
- [41] Campbell LP, Alexander AM. Landscape genetics of *Aedes mcintoshi* (Diptera: Culicidae), an important vector of rift valley fever virus in northeastern Kenya. *Journal of Medical Entomology*. 2017;**54**(5):1258-1265
- [42] Lewin HA, Robinson GE, Kress WJ, Baker WJ, Coddington J, Crandall KA, et al. Earth BioGenome project: Sequencing life for the future of life. *Proceedings of the National Academy of Sciences*. 2018;**115**(17):4325-4333