Agrigenomics Research

An Overview of Recent Agrigenomics Research Publications Featuring Illumina® Technology

TABLE OF CONTENTS

4 Introduction

5 Understanding the genome

Genetic variation and polymorphisms

Agriculture and quantitative traits

11 Implementation of genomic selection in agriculture

Establishing a reference genome

Genotyping

- Genotyping arrays and imputation
- Genotyping by sequencing

Genomic selection in action

24 Beyond selection: using genomics for resource monitoring and management

Fisheries and aquaculture

Disease profiling and pathogen surveillance

Food safety and authenticity

Microbial metagenomics in agriculture

- Plants
- Animals

33 The future of agrigenomics

Gene editing

Advanced breeding strategies

36 Sequencing methods

Genome assembly and epigenetics

- Chipmentation: chromatin immunoprecipitation with sequencing library preparation by tn5 transposase
- Chia-pet: chromatin interaction analysis by paired-end tag sequencing
- 3-c, capture-c, and hi-c: chromatin conformation capture sequencing
- Ng capture-c: next-generation capture-c
- 4c-seq: circular chromatin conformation capture
- Umi-4c: circular chromosome conformation capture with unique molecular identifiers
- 5c: chromatin conformation capture carbon copy

Genotyping by sequencing

- Rad and pe rad-seq: restriction-site associated dna sequencing
- Ddradseq: double digest restriction-site associated dna marker generation
- 2b-rad: rad with type iib restriction endonucleases
- Slaf-seq: specific locus amplified fragment sequencing
- Hyrad: hybridization rad for degraded dna
- Rapture: restriction-site associated dna capture

For Research Use Only. Not for use in diagnostic procedures.

INTRODUCTION

Agriculture, the science and practice of farming plants and animals for food, fiber, and fuel, has sustained and enhanced human life for millennia. Given the changing environment, expanding population, and increasing demands for nutrition, the need to optimize agriculture is of fundamental importance[1]. Harnessing genomic innovations in agriculture will enable more productive and sustainable practices to address these challenges.

Advances in genomic selection (GS) and more recently, gene editing, speed the process of developing crops and livestock with desirable agronomic traits (e.g., higher production value, stress tolerance, disease resistance, and sustainability).[2,3,4] GS is based on the principle that information from a large number of markers distributed across the genome can be used to capture diversity in a population and estimate breeding values of individuals. It was first described in 2001 and hinges on developing a breeding equation using a training population with defined traits.[5] While breeders have used limited sets of molecular markers for decades, advances in genome-wide technologies have led to a dramatic increase in the breadth and depth of genetic resources available for agriculture.

Innovations in genetics, bioinformatics, and biotechnology present breeders with powerful tools to advance agriculture. Sequence data and well-characterized marker sets can be used to correlate genotypes to phenotypes, and the curation of databases containing genomic information are essential for understanding genetic diversity across populations and environments. These data allow us to characterize new species, perform meta-analyses, unravel complex traits, and ultimately empower genomic selection and gene editing for the advancement of agriculture. Genomic technologies have enabled the field of agrigenomics to revolutionize the breeding and management of crops and livestock.

1. FAO The State of Food Security and Nutrition in the World 2018. *Building climate resilience for food security and nutrition*. http://www.fao.org/3/I9553EN/i9553en.pdf
2. Bevan MW, Uauy C, Wulff BBH, Zhou J, Krasileva K and Clark MD Genomic innovation for crop improvement. *Nature*. 2017;543:346-354.
3. Georges M, Charlier C and Hayes B Harnessing genomic information for livestock improvement. *Nat Rev Genet*. 2018;
4. Bernatchez L, Wellenreuther M, Araneda C, et al. Harnessing the Power of Genomics to Secure the Future of Seafood. *Trends Ecol Evol*. 2017;32:665-680.
5. de Koning DJ Meuwissen et al. on Genomic Selection. *Genetics*. 2016;203:5-7.

Figure 1. Agriculture is the keystone of the global economy and health.

UNDERSTANDING THE GENOME

Anything that lives, from the simplest forms of life to more complex organisms, has a genome. Deoxyribonucleic acid (DNA) is the molecule containing genetic information, often described as a blueprint for an organism. DNA encodes and regulates genes, which may be translated into proteins (by way of ribonucleic acid [RNA]). The biological information stored within the DNA genome ultimately orchestrates the development of an organism.

The basic chemical makeup of DNA is the same in all species. It is the combination of two polymeric strands coiled around each other, forming the well-known double-helix structure (Figure 2). Each strand of DNA is composed of a sequence of subunits called nucleotides. Each nucleotide subunit consists of a phosphate group, a monosaccharide, and a variable nitrogen-containing base. DNA contains four possible base subunits: adenine (A), guanine (G), cytosine (C) and thymine (T). The double-helix structure is maintained by the hydrogen bonds between complementry subunits, which are very specific: A pairs with T, and G pairs with C. For this binding to be possible, the two strands are in a reverse and complementary orientation to each other.

Genes, stretches of DNA that encode proteins or regulatory RNAs, are considered the basic unit of heredity in the genome—they shape the traits that characterize an organism.

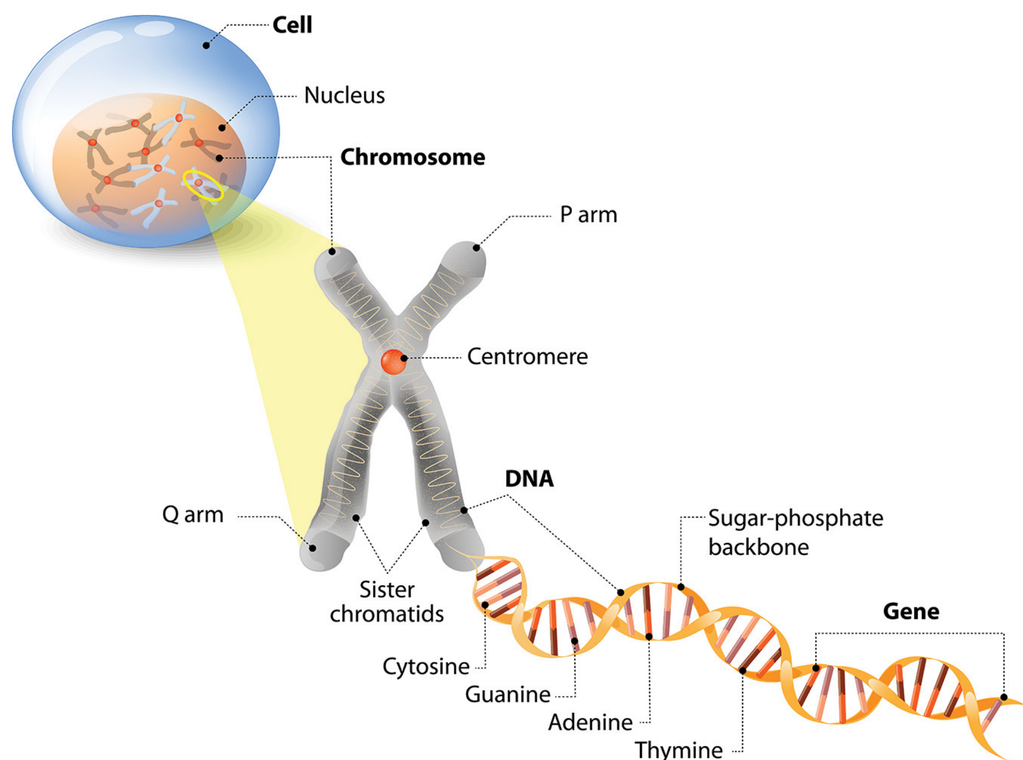


Figure 2. A double helix is formed by two complementary strands of DNA composed of the 4 base nucleotides: A, T, C, and G. Base pairing between nucleotide pairs A—T or C—G is the result of electrostatic interactions called hydrogen bonds. The DNA double helix is wound around proteins and compacted into discrete structures—called chromosomes—which reside within the nucleus of plant and animal cells.

Genetic variation and polymorphisms

Plant and animal genomes typically contain replicate sets of chromosomes—large discrete DNA molecules wound around structural proteins—which are inherited from each parent. Thus, for each gene, diploid organisms (those that contain a duplicate set of chromosomes) will each have two alleles. One allele corresponds to the chromosome that is inherited from the mother, and a second allele corresponds to the chromosome that is inherited from the father. In this way, alleles are alternative forms of the same gene that differ slightly in their nucleotide sequence. The combination of alleles inherited by an individual organism is what defines the genotype of that individual (Figure 3).

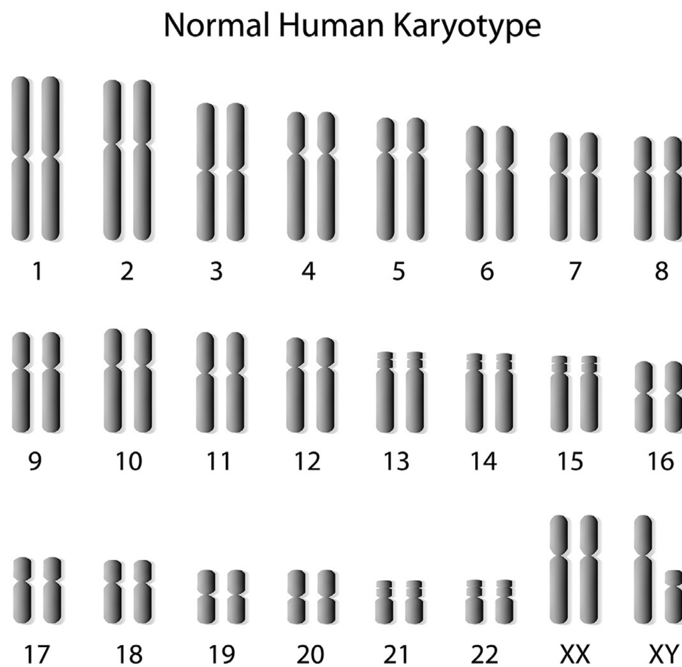


Figure 3. The diploid human karyotype shows the number and relative size of all chromosomes within the nucleus of a cell.

Genetic variability in conjunction with environmental factors accounts for the phenotypic variability seen within and between species. Although much of the genome is conserved for individuals belonging to the same species, there are important positions in which there is inter-individual variability. These positions of DNA variation within genomes of the same species are called polymorphisms. While some polymorphisms are neutral—appearing to have no phenotypic effect—many others account for what differentiates one individual from another. Polymorphisms consist of nucleotide substitutions, insertions, deletions, or repeats, and are categorized based on their sequence and length (Figure 4). Polymorphisms can range from a single nucleotide polymorphism (SNP) to structural variants encompassing tens to thousands of nucleotides. Likewise, the functional effect of any given DNA polymorphism on a phenotype can range from fully penetrant— that is having a strong deterministic effect on phenotype—to having no discernable effect. Most agronomic traits are complex or multifactorial, which leads them to have an intermediate to low penetrance.

Advances in next-generation sequencing (NGS) technologies now allow us to isolate DNA (or RNA) from many types of samples, amplify and sequence regions of the genome, and sequence and assemble whole genomes. This has enabled more robust and widespread identification of polymorphisms that explain phenotypic variability in crops and livestock.

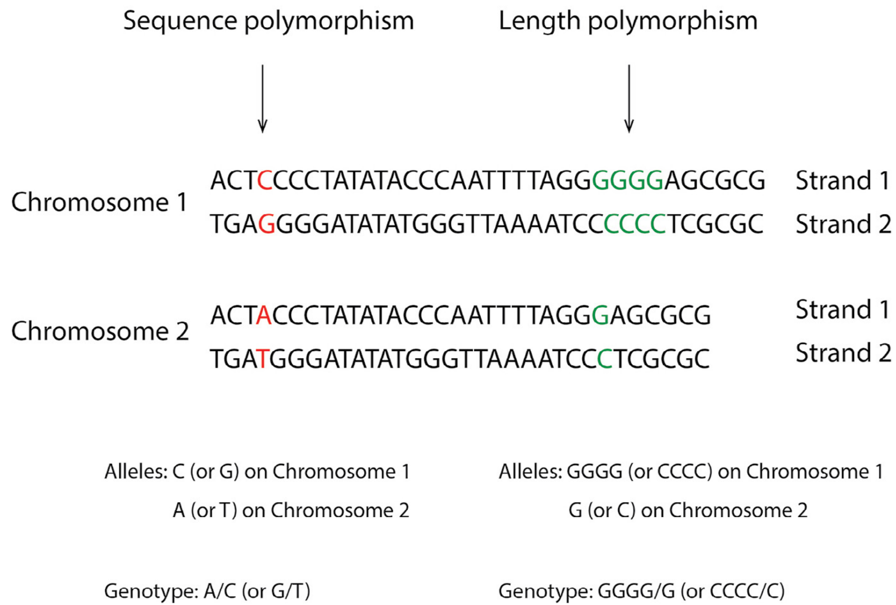


Figure 4. Genetic variation can include DNA sequence or length polymorphisms.

Agriculture and quantitative traits

Both animals and plants evolve in complex environments, acquiring the ability to cope with elements in their environment such as predators, soil conditions, and climate.[6] Most traits, or phenotypes, that are desirable in agriculture are defined as complex (quantitative) traits. Complex traits cannot be explained by the inheritance of a single gene, but rather show continuous variation due to the influence of many genes.[7] In order to understand how genetic variation contributes to complex phenotypic variation, there have been widespread efforts to correlate discrete regions of the genome to a particular trait (phenotype). Identifying these regions, called *quantitative trait loci* (QTL), is critical for identifying and propagating genetic variants that confer beneficial agronomic traits (Table 1).

Historically, selection for breeding has been made using estimated breeding values (EBV), without identifying the genes involved in a particular phenotype. EBVs were estimated from the study of pedigrees and phenotypic records with knowledge of the heritability of each trait (Figure 5). However, the efficiency of this method decreases as it is expanded to traits that are difficult to measure, have low heritability, or can be measured only after several years and/or generations. For this reason, the identification of genes underlying important traits in plants and animals is a major focus of agrigenomics research.

Table 1. Examples of phenotypic traits of agricultural value that are genetically regulated by QTLs, published recently.

Agronomic Trait	Species	References
Yield and growth	Wheat, Rice, Maize, Millet, Potato, Canola, Spinach, Carrot, Sheep, Cattle, Salmon, Yellow Drum	8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25
Disease resistance	Wheat, Rice, Barley, Potato, Sugar beet, Sugarcane, Peanut, Hot pepper, Cattle, Haddock, Catfish	13,17, 22, 25, 26, 27, 28, 29, 30, 31, 32, 33
Abiotic stress adaptation	Wheat, Soybean, Barley, Chickpea, Common bean, Wheat, Millet, Grape, Cattle, Sheep, Goat, Yak, Cod, Catfish	22, 32, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47
Reproduction	Wheat, Millet, Yam, Cattle, Salmon, Yellow Drum, Catfish	12,18, 22, 25, 32, 48, 49, 50
Nutrition/End-use quality	Wheat, Maize, Apple, Potato, Tomato, Olive, Peanut, Carrot, Sheep, Pig, Cattle	13,15,18, 21, 22, 25, 51, 52, 53, 54, 55, 56, 57, 58, 59
Sustainability	Rice, Barley, Chickpea, Cattle, Cod, Tuna	30, 37, 60, 61, 62

6. Anderson JT, Wagner MR, Rushworth CA, Prasad KVSK and Mitchell-Olds T The evolution of quantitative traits in complex environments. *Heredity*. 2014;112:4-12.

7. Boyle EA, Li YI and Pritchard JK An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell*. 2017;169:1177-1186.

8. Yano K, Yamamoto E, Aya K, et al. Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nature Genetics*. 2016;48:927-934.

9. Zhang X, Huang C, Wu D, et al. High-Throughput Phenotyping and QTL Mapping Reveals the Genetic Architecture of Maize Plant Growth. *Plant Physiol*. 2017;173:1554-1564.

10. Romero Navarro JA, Willcox M, Burgueno J, et al. A study of allelic diversity underlying flowering-time adaptation in maize landraces. *Nat Genet*. 2017;49:476-480.

11. Rodgers-Melnick E, Vera DL, Bass HW and Buckler ES Open chromatin reveals the functional maize genome. *Proc Natl Acad Sci USA*. 2016;113:E3177-3184.

12. Pucher A, Hash CT, Wallace JG, Han S, Leiser WL and Haussmann BIG Mapping a male-fertility restoration locus for the A4 cytoplasmic-genic male-sterility system in pearl millet using a genotyping-by-sequencing-based linkage map. *BMC Plant Biol*. 2018;18:65.

13. Hardigan MA, Laimbeer FPE, Newton L, et al. Genome diversity of tuber-bearing Solanum uncovers complex evolutionary history and targets of domestication in the cultivated potato. *Proc Natl Acad Sci USA.* 2017;114:E9999-E10008.

14. Lu K, Xiao Z, Jian H, et al. A combination of genome-wide association and transcriptome analysis reveals candidate genes controlling harvest index-related traits in Brassica napus. *Sci Rep*. 2016;6:36452.

15. Brito LF, Clarke SM, McEwan JC, et al. Prediction of genomic breeding values for growth, carcass and meat quality traits in a multi-breed sheep population using a HD SNP chip. *BMC Genet*. 2017;18:7.

16. Bouwman AC, Daetwyler HD, Chamberlain AJ, et al. Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. *Nat Genet*. 2018;50:362-367.

17. Fang L, Sahana G, Su G, et al. Integrating Sequence-based GWAS and RNA-Seq Provides Novel Insights into the Genetic Basis of Mastitis and Milk Production in Dairy Cattle. *Sci Rep*. 2017;7:45560.

18. Wang T, Chen YP, MacLeod IM, Pryce JE, Goddard ME and Hayes BJ Application of a Bayesian non-linear model hybrid scheme to sequence data for genomic prediction and QTL mapping. *BMC Genomics*. 2017;18:618.

19. Clark EL, Bush SJ, McCulloch MEB, et al. A high resolution atlas of gene expression in the domestic sheep (Ovis aries). *PLoS Genetics*. 2017;13:1-38.

20. Xu C, Jiao C, Sun H, et al. Draft genome of spinach and transcriptome diversity of 120 Spinacia accessions. *Nat Commun*. 2017;8:15275.

21. Iorizzo M, Ellison S, Senalik D, et al. A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. *Nat Genet*. 2016;48:657-666.

22. Appels R, Eversole K, Feuillet C, et al. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*. 2018;361:eaar7191.

23. Waters CD, Hard JJ, Brieuc MSO, et al. Genomewide association analyses of fitness traits in captive-reared Chinook salmon: Applications in evaluating conservation strategies. *Evol Appl*. 2018;11:853-868.

24. Qiu C, Han Z, Li W, Ye K, Xie Y and Wang Z A high-density genetic linkage map and QTL mapping for growth and sex of yellow drum (Nibea albiflora). *Sci Rep*. 2018;8:17271.

25. Naval-Sanchez M, Nguyen Q, McWilliam S, et al. Sheep genome functional annotation reveals proximal regulatory elements contributed to the evolution of modern breeds. *Nat Commun*. 2018;9:859.

26. Stein JC, Yu Y, Copetti D, et al. Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus Oryza. *Nat Genet*. 2018;50:285-296.

27. Capistrano-Gossmann GG, Ries D, Holtgrawe D, et al. Crop wild relative populations of Beta vulgaris allow direct mapping of agronomically important genes. *Nat Commun*. 2017;8:15708.

28. Bertioli DJ, Cannon SB, Froenicke L, et al. The genome sequences of Arachis duranensis and Arachis ipaensis, the diploid ancestors of cultivated peanut. Nat Genet. 2016;48:438-446.

29. Kim S, Park J, Yeom S-i, et al. New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication. *Genome Biology*. 2017;18:210.

30. Milner SG, Jost M, Taketa S, et al. Genebank genomics highlights the diversity of a global barley collection. *Nat Genet*. 2018;

31. Torresen OK, Brieuc MSO, Solbakken MH, et al. Genomic architecture of haddock (Melanogrammus aeglefinus) shows expansions of innate immune genes and short tandem repeats. *BMC Genomics*. 2018;19:240.

32. Kim OTP, Nguyen PT, Shoguchi E, et al. A draft genome of the striped catfish, Pangasianodon hypophthalmus, for comparative analysis of genes relevant to development and a resource for aquaculture improvement. *BMC Genomics*. 2018;19:733.

33. Zhang J, Zhang X, Tang H, et al. Allele-defined genome of the autopolyploid sugarcane Saccharum spontaneum L. *Nat Genet*. 2018;50:1565-1573.

34. Wang T, Chen YPP, MacLeod IM, Pryce JE, Goddard ME and Hayes BJ Application of a Bayesian non-linear model hybrid scheme to sequence data for genomic prediction and QTL mapping. *BMC Genomics*. 2017;18:1-23.

35. Anderson JE, Kono TJ, Stupar RM, Kantar MB and Morrell PL Environmental Association Analyses Identify Candidates for Abiotic Stress Tolerance in Glycine soja, the Wild Progenitor of Cultivated Soybeans. *G3 (Bethesda)*. 2016;6:835-843.

36. Russell J, Mascher M, Dawson IK, et al. Exome sequencing of geographically diverse barley landraces and wild relatives gives insights into environmental adaptation. *Nature Genetics*. 2016;48:1024-1030.

37. Von Wettberg EJB, Chang PL, Basdemir F, et al. Ecology and genomics of an important crop wild relative as a prelude to agricultural innovation. *Nat Commun*. 2018;9:649.

38. Zhou S, Yan B, Li F, et al. RNA-Seq Analysis Provides the First Insights into the Phylogenetic Relationship and Interspecific Variation between Agropyron cristatum and Wheat. *Front Plant Sci*. 2017;8:1644.

39. Varshney RK, Shi C, Thudi M, et al. Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. *Nat Biotechnol*. 2017;35:969-976

40. Xie H, Konate M, Sai N, et al. Global DNA Methylation Patterns Can Play a Role in Defining Terroir in Grapevine (Vitis vinifera cv. Shiraz). *Front Plant Sci*. 2017;8:1860.

41. Yang J, Li WR, Lv FH, et al. Whole-Genome Sequencing of Native Sheep Provides Insights into Rapid Adaptations to Extreme Environments. *Mol Biol Evol*. 2016;33:2576-2592.

42. Song S, Yao N, Yang M, et al. Exome sequencing reveals genetic differentiation due to high-altitude adaptation in the Tibetan cashmere goat (Capra hircus). *BMC Genomics*. 2016;17:122.

43. Zhang X, Wang K, Wang L, et al. Genome-wide patterns of copy number variation in the Chinese yak genome. *BMC Genomics*. 2016;17:379.

44. Berg PR, Star B, Pampoulie C, et al. Three chromosomal rearrangements promote genomic divergence between migratory and stationary ecotypes of Atlantic cod. *Sci Rep*. 2016;6:23246.

45. Barth JMI, Berg PR, Jonsson PR, et al. Genome architecture enables local adaptation of Atlantic cod despite high connectivity. *Molecular Ecology*. 2017;26:4452-4466.

46. Kijewska A, Malachowicz M and Wenne R Alternatively spliced variants in Atlantic cod (Gadus morhua) support response to variable salinity environment. *Sci Rep*. 2018;8:11607.

47. Kim J, Hanotte O, Mwai OA, et al. The genome landscape of indigenous African cattle. *Genome Biol*. 2017;18:34.

48. Charlier C, Li W, Harland C, et al. NGS-based reverse genetic screen for common embryonic lethal mutations compromising fertility in livestock. *Genome Res*. 2016;26:1333-1341.

49. Tamiru M, Natsume S, Takagi H, et al. Genome sequencing of the staple food crop white Guinea yam enables the development of a molecular marker for sex determination. BMC Biol. 2017;15:86.

50. Qiu C, Han Z, Li W, Ye K, Xie Y and Wang Z A high-density genetic linkage map and QTL mapping for growth and sex of yellow drum (Nibea albiflora). *Sci Rep*. 2018;8:17271.

51. Duan N, Bai Y, Sun H, et al. Genome re-sequencing reveals the history of apple and supports a two-stage model for fruit enlargement. *Nat Commun*. 2017;8:249.

For Research Use Only. Not for use in diagnostic procedures.

52. Diepenbrock CH, Kandianis CB, Lipka AE, et al. Novel Loci Underlie Natural Variation in Vitamin E Levels in Maize Grain. *Plant Cell*. 2017;29:2374-2392.

53. Unver T, Wu Z, Sterck L, et al. Genome of wild olive and the evolution of oil biosynthesis. *Proc Natl Acad Sci USA*. 2017;114:E9413-E9422.

54. González-Prendes R, Quintanilla R, Cánovas A, et al. Joint QTL mapping and gene expression analysis identify positional candidate genes influencing pork quality traits. *Scientific Reports*. 2017;7:1-9.

55. Pausch H, Emmerling R, Gredler-Grandl B, Fries R, Daetwyler HD and Goddard ME Meta-analysis of sequence-based association studies across three cattle breeds reveals 25 QTL for fat and protein percentages in milk at nucleotide resolution. *BMC Genomics*. 2017;18:853.

56. Magalhães AFB, Schenkel FS, Garcia DA, et al. Genomic selection for meat quality traits in Nelore cattle. Meat Science. 2019;148:32-37.

57. Zhu G, Wang S, Huang Z, et al. Rewiring of the Fruit Metabolome in Tomato Breeding. Cell. 2018;172:249-261 e212.

58. Tieman D, Zhu G, Jr. MFRR, et al. A chemical genetic roadmap to improved tomato flavor. Science. 2017;355:391-394.

59. Luo H, Pandey MK, Khan AW, et al. Discovery of genomic regions and candidate genes controlling shelling percentage using QTL-seq approach in cultivated peanut (Arachis hypogaea L.). *Plant Biotechnol J*. 2018; Dec 13. doi: 10.1111/pbi.13050.

60. Chung NC, Szyda J and Fraszczak M Population Structure Analysis of Bull Genomes of European and Western Ancestry. *Scientific Reports*. 2017;7:1-10.

61. Wing RA, Purugganan MD and Zhang Q The rice genome revolution: from an ancient grain to Green Super Rice. *Nat Rev Genet*. 2018;19:505-517.

62. Pecoraro C, Babbucci M, Franch R, et al. The population genomics of yellowfin tuna (Thunnus albacares) at global geographic scale challenges current stock delineation. *Sci Rep*. 2018;8:13890.

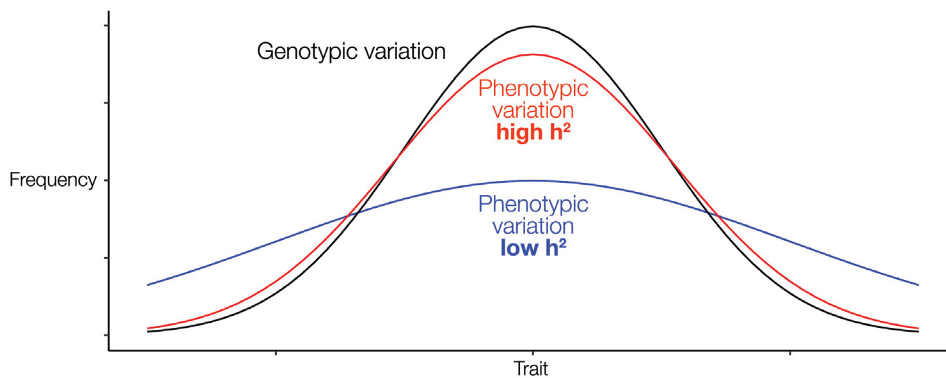


Figure 5. Heritability (often denoted as h^2) is the proportion of a phenotypic trait that can be attributed to genetic factors rather than environment. A small h^2 value is closer to 0 and indicates the trait is strongly influenced by the environment (e.g. yield). A larger h^2 value is closer to 1 and indicates the trait is only slightly influenced by the environment (e.g. flower color).

Available Resources

Animal QTLdb and CorrDB:[63] www.animalgenome.org/QTLdb

Online Mendelian Inheritance in Animals (OMIA):[64] https://omia.org

Gramene QTL Database:[65] https://archive.gramene.org/qtl/

IMPLEMENTATION OF GENOMIC SELECTION IN AGRICULTURE

The implementation of GS for breeding in agriculture has its roots in marker assisted selection (MAS)—a process by which breeders use markers (or QTLs) that are correlated to a trait of interest to select for inheritance of that particular trait indirectly. Historically, QTL studies had limited impact driving MAS for crop and livestock improvement. Early studies suffered from low genetic resolution resulting in sparse marker densities which were not widely applicable across different populations and environments. However, advances in genomic technologies—the development of genome arrays and the advent of NGS platforms—have boosted the ability of researchers to identify genetic markers that explain phenotypic variation in complex traits. The results obtained from recent genome-wide studies across populations and environments have wider utility to drive implementation of GS in crop and livestock breeding programs.

GS is based on the principle that information from a large number of markers can be used to estimate breeding values without having a precise knowledge of where specific genes function or reside. It is similar to conventional MAS in that genetic information is being used, but rather than focusing on a small subset of defined markers (e.g., QTLs) for a single trait (as in MAS), markers throughout the genome can be selected from a training population exhibiting a favorable phenotype. In GS, genomic EBVs (gEBVs) are calculated from the cumulative effect of a large number of genetic markers and then these values are used to score potential breeding candidates.

The successful implementation of GS relies on genetic markers that represent the complete genome, the availability of representative cohorts of individuals belonging to the same species, and genomic prediction algorithms that combine genetic information with phenotypic and/or pedigree data. In contrast to MAS, GS has been widely adopted due to the accessibility of affordable genotyping resources based on NGS and genome arrays. For example, the US dairy cattle industry has largely adopted GS to enhance livestock breeding programs, resulting in more than 3 million animals being genotyped since 2008, decreases in generation intervals for bulls, and rapid improvement of animal health, fertility, and lifespan.[66,67] In addition, GS in conjunction with high-throughput phenotyping (HTP) is gaining traction for accelerating crop breeding programs aimed at increasing grain production.[68,69] GS is proving a powerful tool to improve the efficiency of crop and livestock breeding programs.

63. Hu Z-L, Park CA and Reecy JM Building a livestock genetic and genomic information knowledgebase through integrative developments of Animal QTLdb and CorrDB. *Nucleic Acids Research*. 2018;1-10.
64. *Science SSoV Online Mendelian Inheritance in Animals, OMIA*. http://omia.org/
65. Tello-Ruiz MK, Naithani S, Stein JC, et al. Gramene 2018: unifying comparative genomics and pathway resources for plant research. *Nucleic Acids Research*. 2018;46:D1181-D1189.
66. García-Ruiz A, Cole JB, VanRaden PM, Wiggans GR, Ruiz-López FJ and Van Tassell CP Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. Proceedings of the *National Academy of Sciences*. 2016;113:E3995-E4004.
67. Wiggans GR, Cole JB, Hubbard SM and Sonstegard TS Genomic Selection in Dairy Cattle: The USDA Experience. *Annu Rev Anim Biosci*. 2017;5:309-327.
68. Crossa J, Pérez-Rodríguez P, Cuevas J, et al. Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. *Trends in Plant Science*. 2017;22:961-975.
69. Araus JL, Kefauver SC, Zaman-Allah M, Olsen MS and Cairns JE Translating High-Throughput Phenotyping into Genetic Gain. *Trends Plant Sci*. 2018;23:451-466.

Figure 6. Large-scale GS programs have been implemented for Holstein cattle in the US.

For Research Use Only. Not for use in diagnostic procedures.

Reviews

Bevan MW, Uauy C, Wulff BBH, Zhou J, Krasileva K and Clark MD Genomic innovation for crop improvement. *Nature*. 2017;543:346-354.

Georges M, Charlier C, Hayes B. Harnessing genomic information for livestock improvement. *Nat Rev Genet*. 2019;20(3):135-156.

Brozynska M, Furtado A and Henry RJ Genomics of crop wild relatives: expanding the gene pool for crop improvement. *Plant Biotechnol* J. 2016;14:1070-1085.

Establishing a reference genome

To date, hundreds of species have been sequenced at high coverage, and these data are available in public databases. However, a *de novo* sequencing project is necessitated when a reference genome is unavailable or if the current reference is of insufficient quality. The substantial efforts invested in nucleotide sequencing technologies and the introduction of NGS have resulted in a 100 million-fold decrease in the cost per genome since 1990. To date, the cost for a raw megabase (Mb) of DNA sequence is less than $0.1, and *de novo* sequencing projects have become routine (Table 2).

Plant and animal genomes tend to vary widely in their size and proportion of repetitive sequences. With the exception of certain species of fish, livestock animal genomes are diploid. Plants often have much more complex genomes that contain multiple sets of homologous chromosomes—termed polyploidy. In order to better characterize complex polyploid genomes, researchers may construct a pan-genome—the result of sequencing individuals from multiple closely related species (Figure 7).

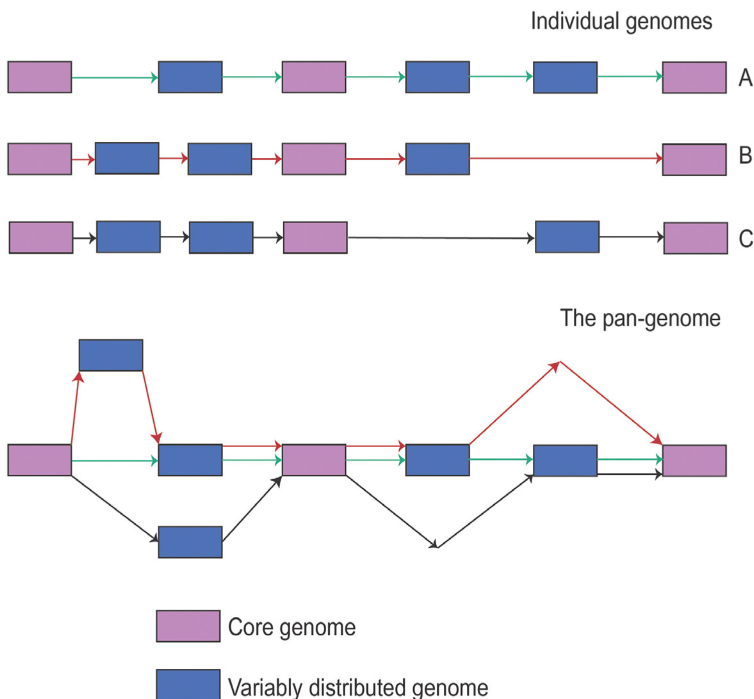


Figure 7. The pan-genome summarizes information about structural variation across the genomes of closely related species or individuals. It includes a core genome that is present in all sequences and a variable distributed genome that represents sequence variations.

Alternately, researchers may deconstruct the genome of a polyploid crop into its sub-genomes. In addition to multiple sets of homologous chromosomes, plant genomes tend to contain higher percentages of repetitive sequences, which make them more difficult to reconstruct with high accuracy. Similarly, determining the ancestral origin of plant and animal genome structural variants also requires specialized tools, which can resolve haplotypes—the set of alleles that are likely to be inherited together (Figure 8).

Sequencing of reference population (Sequence length = 51 bases)

Identification of haplotypes, and design array that uses a few SNPs to represent the region

Figure 8. Haplotypes are groups of multiple alleles that are likely to be inherited together because of linkage disequilibrium. They can be identified through sequencing of multiple individuals and can provide a set of representative SNPs for genotyping.

Recent advances in genomic DNA library preparation, assays of chromosome architecture, and genome assembly algorithms have empowered the generation of more accurate and contiguous reference genomes while maintaining a reasonable cost (Table 2). Highly contiguous, high-quality reference genomes facilitate gene annotation, are important for understanding chromosome evolution and genetic variation, and are essential for the practical application of gene editing technologies.

Reviews

Eagen KP Principles of Chromosome Architecture Revealed by Hi-C. Trends Biochem Sci. 2018;43:469-478.

Rice ES and Green RE New Approaches for Genome Assembly and Scaffolding. Annu Rev Anim Biosci. 2019;7:2018; Nov 28.

Jiao WB and Schneeberger K The impact of third generation genomic technologies on plant genome assembly. Curr Opin Plant Biol. 2017;36:64-70.

Kyriakidou M, Tai HH, Anglin NL, Ellis D and Strömvik MV Current Strategies of Polyploid Plant Genome Sequence Assembly. Front Plant Sci. 2018;9:

Phillippy AM New advances in sequence assembly. Genome Research. 2017;27:xi-xiii.

Table 2. Select reference genome assemblies for crop and livestock species, published recently.

Species	Genome Size	Scaffold/contig N50	Sequencing Platforms	Ref
Z. mays (Maize B73)	2.3 Gb	89 kb/14.5 kb	HiSeq X Ten	70
Z. mays (Maize W22)	2.2 Gb	35.5 Mb/72 kb	HiSeq 2500, HiSeq 2000, HiSeq X Ten	71
T. aestivum L. (Wheat)	14.5 Gb	7 Mb/52 kb	HiSeq 2500, NextSeq 500, PacBio RS II	22
T. urartu (Wheat A subgenome)	4.9 Gb	3.67 Mb/344 kb	HiSeq 2500, PacBio RS II	72
C. americanus L. (Pearl Millet)	1.8 Gb	885 kb/18 kb	HiSeq 2000	73
P. miliaceum (Proso Millet)	1.0 Gb	912 kb/	HiSeq X Ten	70
C. annum (Chili Pepper)	3.5 Gb	3.69 Mb/123 kb	HiSeq X Ten	74
C. baccatum (Chili Pepper)	3.9 Gb	2.0 Mb/39 kb	HiSeq 2500	29
C. Chinese (Chili Pepper)	3.2 Gb	3.3 Mb/50 kb	HiSeq 2500	29
P. somniferum L. (Opium Poppy)	2.7 Gb	204 Mb/1.77 Mb	HiSeq 2500, HiSeq 4000, PacBio RS II	75
D. carota (Carrot)	473 Mb	12.7 Mb/31.2 kb	HiSeq 2000	21
C. quinoa (Quinoa)	1.5 Gb	3.84 Mb/1.66 Mb	HiSeq 2000, PacBio RS II	76
H. vulgare L. (Barley)	4.8 Gb	1.9 Mb/79 kb	MiSeq, HiSeq 2000, HiSeq 2500	77
D. rotundata (Yam)	594 Mb	2.12 Mb/	HiSeq 2500, MiSeq	49
S. spontaneum L. (Sugarcane)	3.4 Gb	/45 kb	HiSeq 2500, PacBio RS II	33
S. oleracea (Spinach)	1.0 Gb	919 kb/16.5 kb	HiSeq 2500	20
A. officinalis L. (Asparagus)	1.3 Gb	301 kb/	NextSeq 500	78
C. hircus (Goat)	2.9 Gb	87.3 Mb/ 18.7 Mb	HiSeq 2500, MiSeq, PacBio RS II	79
G. gallus (Chicken)	1.2 Gb	6.4 Mb/2.9Mb	HiSeq 2000, PacBio RS II	80
S. salar (Atlantic Salmon)	2.2 Gb	57.6 kb/ 2.97 Mb	HiSeq	81
G. morhua (Atlantic Cod)	613 Mb	116 kb/1.15 Mb	HiSeq 2000, PacBio RS	82
M. aeglefinus (Haddock)	644 Mb	116 kb/1.15 Mb	HiSeq 2000, PacBio RS	31
P. hypophthalmus (Striped Catfish)	1.0 Gb	6 kb/8.21 Mb	MiSeq, HiSeq 2500	32
P. yessoensis (Scallop)	1.43 Gb	38 kb/804 kb	HiSeq 2000	83

Cracking the genetic code of bread wheat

Bread wheat is one of the world's most widely cultivated crops. A new high-quality annotated reference genome, published by the International Wheat Genome Sequencing Consortium (IWGSC), will be a valuable resource in the effort to accelerate improvement of this staple crop.[84]

“The wheat genome sequence lets us look inside the wheat engine…What we see is beautifully put-together to allow for variation and adaptation to different environments through selection, as well as sufficient stability to maintain basic structures for survival under various climatic conditions.” -IWGSC

Figure 9. Wheat has one of the most complex genomes known.

References

Appels R, Eversole K, Feuillet C, et al. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*. 2018;361:eaar7191.

Bread wheat has a complex genome containing three sets of homologous chromosomes and more than 5 times the genetic information as the human genome. The International Wheat Genome Sequencing Consortium (IWGSC) recently released a fully annotated, chromosome-scale assembly of the hexaploid bread wheat genome. The 21 annotated pseudomolecules comprising the A, B and D subgenomes are estimated to contain over 100,000 high-confidence gene models and achieve 94% coverage. The assembly enabled subsequent discovery of tissue and developmental gene coexpression networks from a transcriptome atlas and revealed dynamics of complex gene family evolution in response to environmental adaptation and end-use quality. These data were then compared to known agronomic markers and QTLs. Ultimately, this reference assembly demonstrates its utility for wheat improvement through the discovery of the genetic basis of biotic and abiotic stress adaptation, and by informing a genome editing strategy for a trait impacting flowering-time.

Illumina technology: HiSeq 2500 System, NextSeq 500 System

Bickhart DM, Rosen BD, Koren S, et al. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nature Genetics*. 2017;49:643-650.

Decreasing sequencing costs have led to an explosion of sequencing whole genomes to generate new reference assemblies. Despite new advances in short read sequence assembly algorithms, many of these references remain fragmented, with many gaps and structural errors. Using a combination of short and long read sequencing technology, chromosome conformation capture (Hi-C) and optical interaction mapping, a new high-quality reference genome for the domesticated goat (C. hircus) was generated. The resulting assembly is the most continuous de novo mammalian assembly produced to date. Chromosome level scaffolds, minimal gaps and highly resolved repetitive regions represent a huge improvement over the previous assembly. This continuous, accurate reference assembly enables breeders access to a vital resource for advanced genomic selection and gene editing approaches.

Illumina technology: HiSeq System, MiSeq System, Caprine53K SNP BeadChip

70. Ott A, Schnable JC, Yeh C-t, et al. Linked read technology for assembling large complex and polyploid genomes. *BMC Genomics*. 2018;19:651.
71. Springer NM, Anderson SN, Andorf CM, et al. The maize W22 genome provides a foundation for functional genomics and transposon biology. *Nature Genetics*. 2018;50:1282-1288
72. Ling H-q, Ma B, Shi X, et al. Genome sequence of the progenitor of wheat A subgenome Triticum urartu. *Nature*. 2018;557:424-428.
73. Varshney RK, Shi C, Thudi M, et al. Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. *Nat Biotechnol*. 2017;35:969-976
74. Hulse-Kemp AM, Maheshwari S, Stoffel K, et al. Reference quality assembly of the 3.5-Gb genome of Capsicum annuum from a single linked-read library. *Hortic Res*. 2018;5:4.
75. Guo L, Winzer T, Yang X, et al. The opium poppy genome and morphinan production. *Science*. 2018;362:343-347.
76. Jarvis DE, Ho YS, Lightfoot DJ, et al. The genome of Chenopodium quinoa. *Nature*. 2017;542:307-312.
77. Mascher M, Gundlach H, Himmelbach A, et al. A chromosome conformation capture ordered sequence of the barley genome. *Nature*. 2017;544:427-433.
78. Harkess A, Zhou J, Xu C, et al. The asparagus genome sheds light on the origin and evolution of a young Y chromosome. *Nat Commun*. 2017;8:1279.
79. Bickhart DM, Rosen BD, Koren S, et al. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nature Genetics*. 2017;49:643-650.
80. Warren WC, Hillier LW, Tomlinson C, et al. A New Chicken Genome Assembly Provides Insight into Avian Genome Structure. *G3 (Bethesda)*. 2017;7:109-117.
81. Lien S, Koop BF, Sandve SR, et al. The Atlantic salmon genome provides insights into rediploidization. *Nature*. 2016;533:200-205.
82. Torresen OK, Star B, Jentoft S, et al. An improved genome assembly uncovers prolific tandem repeats in Atlantic cod. *BMC Genomics*. 2017;18:95.
83. Wang S, Zhang J, Jiao W, et al. Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nat Ecol Evol*. 2017;1:120.
84. *IWGSC The Wheat Code is Finally Cracked. 2018*; https://www.wheatgenome.org/News/Press-releases/The-Wheat-Code-is-Finally-Cracked.

For Research Use Only. Not for use in diagnostic procedures.

Koren S, Rhie A, Walenz BP, et al. De novo assembly of haplotype-resolved genomes with trio binning. *Nat Biotechnol*. 2018;36:12.

Resolving allelic variation during diploid genome assembly is important for constructing haplotype-aware reference genomes. Mosaic consensus genome assemblies can contain false variants (not present in parental haplotypes), which impact downstream analyses. Sequencing inbred individuals circumvents these problems, but this approach is impractical for many species and is not useful for understanding genetic variation within a species or outbred populations. The development of the trio binning approach simplifies haplotype assembly and performs well with high levels of heterozygosity. The approach uses short-read genome sequencing data from two parental genomes to partition an offspring's long-read genome sequencing data into sets by haplotype. Then each haplotype is assembled independently resulting in a haplotype aware diploid genome assembly. The approach has been validated on a F1 cross between cattle subspecies *Bos taurus taurus* and *Bos taurus indicus*. The resulting complete parental haplotype assemblies had haplotig sizes >20 Mb with 99.998% accuracy. This approach significantly improves diploid genome assembly and is informative for studies of genetic variation and inheritance.

Illumina technology: NextSeq 500 System, BovineHD BeadChip

Wu S, Lau KH, Cao Q, Hamilton JP, Sun H, Zhou C, et al. Genome sequences of two diploid wild relatives of cultivated sweetpotato reveal targets for genetic improvement. *Nat Commun*. 2018;9:1-12.

Sweet potatoes are a clonally propagated, heterozygous polyploidy crop that serve as an important staple food source particularly for sub-Saharan Africa. A lack of genomic resources for improvement have limited agronomic gains. Recent high-quality genome assemblies of diploid sweet potato crop wild relatives, *I. trifida* and *I. triloba* provide insights into the whole-genome triplication during *Ipomoea* evolution. Genome resequencing of 16 genotypes commonly used in African breeding programs enabled the identification of genes and alleles associated with carotenoid biosynthesis in storage roots. These genomic resources can serve as the basis for efficient breeding strategies to improve nutrition content in an important staple crop (e.g. breeding varieties with higher provitamin A).

Illumina technology: HiSeq 2500 System

Available Resources

NCBI Reference Genomes: https://www.ncbi.nlm.nih.gov/genome

The Genome 10K project: https://genome10k.soe.ucsc.edu/

Vertebrate Genomes Project: https://vertebrategenomesproject.org/

The Genome OnLine Database (GOLD):[85] https://gold.jgi.doe.gov/index

International Wheat Genome Sequencing Consortium (IWGSC): www.wheatgenome.org

Genotyping

Once the effort of assembling a genome is complete, it is important to find an efficient method to capture the genetic diversity of a species. Typically this is achieved by generating data on informative cohorts of individuals from the species or population of interest. The sequencing and comparison of multiple individuals from a population is an extensive source of genomic markers. Once a reference sequence is available, the assembly of multiple genomes based on their alignment to the reference is a feasible task. Though the cost of whole-genome sequencing has declined significantly in the last decades, it is still prohibitive to deeply sequence the whole genome of large populations in a study or breeding program.

Genotyping Arrays and Imputation

Genotyping arrays (also called microarrays) contain oligonucleotide probes that detect up to hundreds of thousands of genomic markers in parallel (Figure 10). Arrays are designed to select representative polymorphisms (e.g., SNPs) that characterize the genetic variation present in a population. In order to deveop a microarray platform for genotyping, researchers first sequence a representative diversity set of individuals to identify haplotypes (Figure 8).

85. Mukherjee S, Stamatis D, Bertsch J, et al. Genomes OnLine database (GOLD) v.7: updates and new features. *Nucleic Acids Research*. 2019;47:D649-D659

86. *Illumina Nellore Genotyping Allows Brazilian Beef Industry to Flourish*. 2016; https://www.illumina.com/science/customer-stories/icommunity-customer-interviews-case-studies/lobo-ancp-interview-nellore-genotyping.html.

87. Mahmoud M, Zeng Y, Shirali M, et al. Genome-wide pleiotropy and shared biological pathways for resistance to bovine pathogens. *PloS One*. 2018;13:e0194374.

88. Magalhães AFB, Schenkel FS, Garcia DA, et al. Genomic selection for meat quality traits in Nelore cattle. *Meat Science*. 2019;148:32-37.

89. He J, Guo Y, Xu J, et al. Comparing SNP panels and statistical methods for estimating genomic breed composition of individual animals in ten cattle breeds. *BMC Genet*. 2018;19:56.

90. Fang L, Sahana G, Ma P, et al. Use of biological priors enhances understanding of genetic architecture and genomic prediction of complex traits within and between dairy cattle breeds. *BMC Genomics*. 2017;18:604.

91. Koren S, Rhie A, Walenz BP, et al. De novo assembly of haplotype-resolved genomes with trio binning. *Nat Biotechnol*. 2018; 36:12.

92. Zhou Y, Utsunomiya YT, Xu L, et al. Genome-wide CNV analysis reveals variants associated with growth traits in Bos indicus. *BMC Genomics*. 2016;17:419.

93. Van Son M, Enger EG, Grove H, et al. Genome-wide association study confirm major QTL for backfat fatty acid composition on SSC14 in Duroc pigs. *BMC Genomics*. 2017;18:369.

94. Li M, Chen L, Tian S, et al. Comprehensive variation discovery and recovery of missing sequence in the pig genome using multiple de novo assemblies. *Genome Res*. 2017;27:865-874.

95. Yang T, Wilkinson J, Wang Z, Ladinig A, Harding J and Plastow G A genome-wide association study of fetal response to type 2 porcine reproductive and respiratory syndrome virus challenge. *Sci Rep*. 2016;6:20305.

96. Hendricks WPD, Zismann V, Sivaprakasam K, et al. Somatic inactivating PTPRJ mutations and dysregulated pathways identified in canine malignant melanoma by integrated comparative genomic analysis. *PloS Genetics*. 2018;14:e1007589.

97. Jones D, Zenger K, Khatkar M, et al. *Development of a low-density commercial genotyping array for the white legged shrimp, Litopenaeus vannamei*. 2017;

98. Robledo D, Matika O, Hamilton A and Houston RD Genome-Wide Association and Genomic Selection for Resistance to Amoebic Gill Disease in Atlantic Salmon. *G3 (Bethesda, Md.)*. 2018;8:1195-1203.

Due to their low costs, simplified analysis, and strong parallel-processing and automation capabilities, the use of DNA arrays and whole-genome SNP imputation has enabled efficient genotyping of large cohorts of individuals.

Figure 10. The Illumina Infinium® XT microarray is an example of an array platform containing multiple oligonucleotide probes. Complementary DNA (cDNA) hybridizes to these probes enabling detection of DNA polymorphisms for genotyping.

Imputation relies on a reference database of fully sequenced genomes to predict genotypes that are not assayed in a larger sample of individuals. The approach consists of first reconstructing haplotypes for the samples of interest using haplotypes from the reference set (haplotype phasing) and then estimating genotypes (Figure 11).

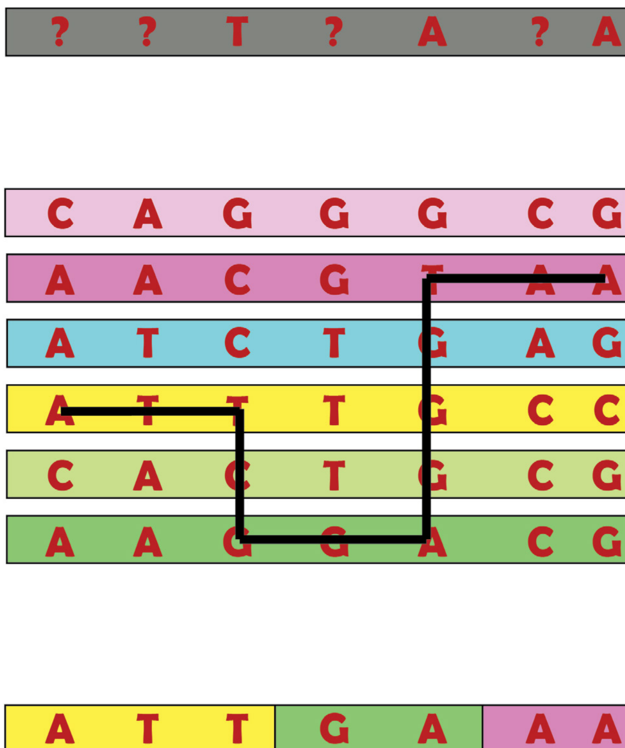


Figure 11. The process of imputation estimates the missing genotypes of an individual by first estimating the haplotypes (called haplotype phasing) from the available genotype data.

99. Liu N, Xue Y, Guo Z, Li W and Tang J Genome-Wide Association Study Identifies Candidate Genes for Starch Content Regulation in Maize Kernels. *Frontiers in Plant Science*. 2016;7:

100. Melchinger AE, Schopp P, Müller D, et al. Safeguarding Our Genetic Resources with Libraries of Doubled-Haploid Lines. *Genetics*. 2017;206:1611-1619.

101. Anderson JE, Kono TJ, Stupar RM, Kantar MB and Morrell PL Environmental Association Analyses Identify Candidates for Abiotic Stress Tolerance in Glycine soja, the Wild Progenitor of Cultivated Soybeans. *G3 (Bethesda)*. 2016;6:835-843.

102. Bayer MM, Rapazote-Flores P, Ganal M, et al. Development and Evaluation of a Barley 50k iSelect SNP Array. *Front Plant Sci*. 2017;8:1792.

103. Clarke WE, Higgins EE, Plieske J, et al. A high-density SNP genotyping array for Brassica napus and its ancestral diploid species based on optimised selection of single-locus markers in the allotetraploid genome. TAG. Theoretical and applied genetics. *Theoretische und angewandte Genetik*. 2016;129:1887-1899.

104. Lu K, Xiao Z, Jian H, et al. A combination of genome-wide association and transcriptome analysis reveals candidate genes controlling harvest index-related traits in Brassica napus. *Sci Rep*. 2016;6:36452.

105. Faivre-Rampant P, Zaina G, Jorge V, et al. New resources for genetic studies in Populus nigra: genome-wide SNP discovery and development of a 12k Infinium array. *Mol Ecol Resour*. 2016;16:1023-1036.

106. Livaja M, Unterseer S, Erath W, et al. Diversity analysis and genomic prediction of Sclerotinia resistance in sunflower using a new 25 K SNP genotyping array. TAG. Theoretical and applied genetics. *Theoretische und angewandte Genetik*. 2016;129:317-329.

107. Hulse-Kemp AM, Ashrafi H, Plieske J, et al. A HapMap leads to a Capsicum annuum SNP infinium array: a new tool for pepper breeding. *Hortic Res*. 2016;3:16036.

The use of arrays and imputation is well established in human, animal, and plant studies. Several platforms have been designed specifically for agriculture applications (Table 3).

Reviews

Berry DP, Garcia JF and Garrick DJ Development and implementation of genomic predictions in beef cattle. *Animal Frontiers*. 2016;6:32-38.

Rasheed A, Hao Y, Xia X, et al. Crop Breeding Chips and Genotyping Platforms: Progress, Challenges, and Perspectives. *Mol Plant*. 2017;10:1047-1064.

Table 3. Illumina commercial, collaborative, and consortia microarray genotyping platforms for crop and livestock species cited in recent publications.

Species	Platform	References
Cattle	BovineHD BeadChip, BovineLD BeadChip v2.0, BovineSNP50 v3 DNA Analysis BeadChip	18,16, 86, 87, 88, 89, 90, 91, 92
Pig	PorcineSNP60 BeadChip v2	93, 94, 95
Sheep	OvineSNP50 BeadChip, OvineHD BeadChip	15
Goat	Caprine53K SNP beadchip	79
Dog	CanineHD Whole-Genome Genotypong BeadChip	96
Shrimp	Infinium ShrimpLD-24 BeadChip v1.0	97
Salmon	Infinium SNP17K BeadChip	98
Cod	Infinium SNP12K BeadChip	45
Maize	MaizeLD BeadChip, MaizeSNP50 BeadChip	99,100
Soy	Infinium SNP50K BeadChip	101
Barley	iSelect SNP50K BeadChip	102
Canola	Infinium SNP60K BeadChip	103,104
Poplar	Infinium SNP12K BeadChip	105
Sunflower	Infinium SNP25K BeadChip	106
Pepper	Infinium 16K BeadChip	107

High-Throughput Genotyping Informs Selective Breeding of Livestock

Neogen Corporation's GeneSeek subsidiary develops custom arrays based on Infinium technology, enabling cost-effective genotyping that provides ranchers with the information they need to improve herds.[108]

For Research Use Only. Not for use in diagnostic procedures.

"Low-cost, high-throughput, high-capacity next-generation sequencing (NGS) technologies have increased the amount of plant and animal DNA sequencing data. [These] data have informed the development of high-quality, predictive fixed arrays that support high-volume, less expensive population screening for genetic variation."

-J. Walker (Neogen/GeenSeek)

Genotyping by Sequencing

Genotyping by sequencing (GBS) is a set of sequencing-based methods for genotyping. GBS methods employ one of two strategies: whole genome resequencing or reduced-representation sequencing. The first GBS strategy uses whole-genome libraries in conjunction with low-coverage "skim" sequencing.[109] A second GBS strategy reduces the complexity of the genome by generating reduced representation libraries (RRLs). These libraries are often constructed by fragmenting the genome through the use of restriction enzymes (REs). Barcoded adapters may also be used to target specific regions of the genome and allow the study of multiple individuals at the same time in a process called multiplexing. The resulting RRLs are then used for high-coverage sequencing. Since this method of GBS does not require *a priori* knowledge of the sequence being studied, it has advantages over array-based methods in cases where information about the spacing of markers along chromosomes is unavailable. GBS is widely used in species where a reference genome or genotyping arrays are unavailable.

GBS capitalizes on the advantages of NGS to quickly and cost-effectively scan genomes of species with or without a reference genome assembly.[110,111]

Reviews

Scheben A, Batley J and Edwards D Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. *Plant Biotechnol J*. 2017;15:149-161.

Xu X and Bai G Whole-genome resequencing: changing the paradigms of SNP detection, molecular mapping and gene discovery. *Molecular Breeding*. 2015;35:

Andrews KR, Good JM, Miller MR, Luikart G and Hohenlohe PA Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat Rev Genet*. 2016;17:81-92.

References

Malmberg MM, Barbulescu DM, Drayton MC, Shinozuka M, et al. Evaluation and Recommendations for Routine Genotyping Using Skim Whole Genome Re-sequencing in Canola. *Frontiers in Plant Science*. 2018;9:1-15.

The decreasing cost of sequencing is enabling genotyping by skim whole genome re-sequencing (skim WGR). A set of 149 samples, representative of global Canola diversity, were analyzed in silico for the application of skim WGR without the use of imputation. Skim samples (0.25x, 0.5x, 1x, 2x, 3x, 4x, and 5x) were generated from raw sequence files (FASTQ) averaging 10x coverage using pre-defined versus de novo SNP discovery. Genotyping pre-defined SNPs at 1-2x coverage and stringent depth filtering was sufficient to genotype the diversity set for Canola. Optimal coverage for skim WGR in a simpler genetic background of double haploid lines was also assessed by down sampling low coverage (1-2x) whole-genome sequencing data. Ultimately, selection of appropriate parameters for skim WGR is dependent on the demands between marker number, accuracy and sequencing cost, which will depend on the application.

Illumina technology: HiSeq 3000 System

108. *Illumina High-Throughput Genotyping Informs Selective Breeding of Livestock. Illumina iCommunity Newsletter*. 2017; https://www.illumina.com/content/dam/illumina-marketing/documents/webicommunity/walker-geneseek-interview-infinium-ag-1370-2017-001.pdf.

109. Huang, X., Feng, Q., Qian, Q., Zhao, Q., et al. High-throughput genotyping by whole-genome resequencing. *Genome Research*. 2009;19: 1068–1076.

110. Torkamaneh D, Laroche J, Bastien M, Abed A and Belzile F Fast-GBS: a new pipeline for the efficient and highly accurate calling of SNPs from genotyping-by-sequencing data. *BMC Bioinformatics*. 2017;18:5.

111. Melo AT, Bartaula R and Hale I GBS-SNP-CROP: a reference-optional pipeline for SNP discovery and plant germplasm characterization using variable length, paired-end genotyping-by-sequencing data. *BMC Bioinformatics*. 2016;17:29.

For Research Use Only. Not for use in diagnostic procedures.

Malmberg MM, Pembleton LW, Baillie RC, Drayton MC, et al. Genotyping-by-sequencing through transcriptomics: implementation in a range of crop species with varying reproductive habits and ploidy levels. *Plant Biotechnology Journal.* 2018;16: 877–889.

Genomic selection is driving rapid improvement of crop and livstock species, yet adoption for many minor crop species is hindered by lack of accessible and cost-effective genotyping strategies. Variations in genome size, ploidy, SNP frequency and modes of reproduction also complicate application of genotyping methods across diverse species. This study describes the use of a genotyping-by-sequencing (GBS) approach based on low-coverage (skim) transcriptome resequencing (GBS-transcriptomics; GBS-t) across a diverse range of crop species (i.e. perennial ryegrass, a diploid outbreeding forage grass; phalaris, a putative segmental allotetraploid outbreeding forage grass; lentil, a diploid inbreeding grain legume; and canola, an allotetraploid partially outbreeding oilseed). With careful consideration for the bioinformatics approcah, GBS-t genotyping proves to be a cost-effective means of SNP genotyping a variety of crop species, including those with large complex genomes or limited genomic reference information.

Illumina technology: NextSeq 500 System, HiSeq 2000/3000 Systems

Genotyping by Sequencing Enables Selective Breeding of Goat Herd

A New Zealand research group uses GBS to overcome cost and resource hurdles associated with genomic selection for orphan crops and minor livestock species.[112]

“GBS makes genomic selection possible and cost effective for any crop or species that someone wants to improve.”

-J. McEwan (AgResearch)

Genomic selection in action

Harnessing GS for crop and livestock selection depends on discovering the genetic basis of fitness and adaptation. Thus, implementation depends on the existence of breeding programs in which genomic, phenotypic, and pedigree information are collected and recorded. Breeders then use this information to select and breed the most promising animals. The advantages of GS are that it can be implemented very early in life (therefore shortening the generation intervals), it is not limited to a single sex, and it is of particular usefulness for traits that are not easy to improve.

The process of GS includes the following steps, also outlined in Figure 12:

- Collection and documentation of phenotype and genotype data for each marker of interest in the reference sample (or discovery dataset).
- Representation of each genotype by a variable, x, that can have 3 values: 0 (homozygote for one allele), 1 (heterozygote), and 2 (homozygote for the second allele).
- Statistical analysis on a reference population to estimate the effect of each marker (w) on the phenotype.
- Generation of a prediction equation for the gEBV that combines all the marker genotypes with their effects on the predictive value of each animal (see below).
- Application of the prediction equation to a group of animals for which genotypes (but not phenotypes) are available. Breeding values are estimated and the best animals are selected for breeding.

112. *Illumina Genotyping by Sequencing Enables Selective Breeding of Goat Herd*. 2017; https://www.illumina.com/science/customer-stories/icommunity-customer-interviews-case-studies/gbs-enables-selective-breeding-of-goat-herd.html.

113. *Falconer D and Mackay T Introduction to Quantitative Genetics*. 1996;

The prediction equation for the gEBV is under constant refinement as breeding programs progress and results are collected. Its estimation from genomic markers can be summarized in the following 3 steps:

1. Use of the markers to deduce the genotype of each animal/plant at each QTL
2. Estimation of the effects of each QTL genotype on the trait
3. Sum of all the QTL effects to finally obtain the gEBV for each individual and select candidates

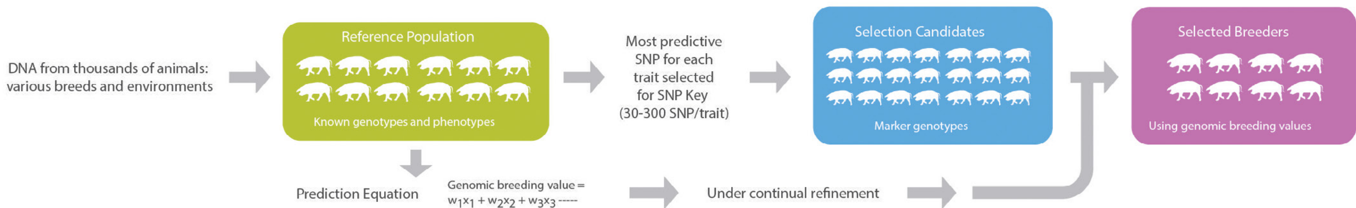


Figure 12. The implementation of genomic selection.

Consider the following equation for genetic gain:[113]

$$\Delta G = \frac{i\ r\ \sigma_A}{L}$$

ΔG, genetic gain; i, selection intensity (directly proportional to population size and proportion selected); r, accuracy (proportional to the reference population size); σ_A, genetic variation within a population; *L*, generational interval

The use of GS allows the breeder to:

- Increase i through the use of larger breeding populations
- Increase r through the use of larger reference populations
- Increase σ_A due to a more robust quantification of variation
- Decrease *L* due to the reduction of the time required to obtain individuals carrying the trait or traits of interest, as shown in Figure 12

All these factors contribute to a higher genetic gain and, consequently, a better yield of an individual carrying the desired qualities. In this way, breeders can use GS to accelerate their breeding programs to develop more productive and sustainable crops and livestock.

Reviews

Meuwissen T, Hayes B and Goddard M Genomic selection: A paradigm shift in animal breeding. *Animal Frontiers*. 2016;6:6-14.

Hickey JM, Chiurugwi T, Mackay I, et al. Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nature Genetics*. 2017;49:1297-1303.

Crossa J, Perez-Rodriguez P, Cuevas J, et al. Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. *Trends Plant Sci*. 2017;22:961-975.

Cerrudo D, Cao S, Yuan Y, et al. Genomic Selection Outperforms Marker Assisted Selection for Grain Yield and Physiological Traits in a Maize Doubled Haploid Population Across Water Treatments. *Front Plant Sci*. 2018;9:366.

Xu Y, Li P, Zou C, et al. Enhancing genetic gain in the era of molecular breeding. *J Exp Bot*. 2017;68: 2641-2666.

For Research Use Only. Not for use in diagnostic procedures.

References

Tieman D, Zhu G, Resende MFR, Lin T, et al. A chemical genetic roadmap to improved tomato flavor. *Science*. 2017;394:391-394.

Commercialized tomato varieties lack the flavor appeal of older heirloom varieties. To understand the genetic basis of tomato flavor, nearly 400 accessions, including commercial, heirloom and wild varieties, were genotyped using whole genome sequencing. In conjunction, a subset of these varieties was evaluated by consumers for flavor, enabling identification of chemicals that contribute to a desirable taste. Modern commercial varieties tend to have less of the chemicals that confer good flavor and genome-wide association study identified genetic loci responsible for important flavor chemicals like sugars, acids and aromatics. This study provides an important resource for commercial tomato improvement through the identification of the genetic basis of flavor traits in tomatoes.

Illumina technology: HiSeq 2000 System

Hardigan MA, Laimbeer FPE, Newton L, Crisovan E, et al. Genome diversity of tuber-bearing Solanum uncovers complex evolutionary history and targets of domestication in the cultivated potato. *Proc Natl Acad Sci*. 2017;114:46.

Cultivated potatoes, domesticated from wild *Solanum* species native to South America have a very diverse gene pool with over 100 varieties. Genotypes from 67 cultivated potato varieties, their progenitors, and out groups were used to capture genome variation and conservation in Solanum species. Roughly 2600 genes were identified as under selection with only modest overlap between North and South American varieties suggesting that a limited gene set drove early improvement of potato cultivars. Genetic signatures of selection were identified in genes regulating carbohydrate metabolism, glycoalkaloid biosynthesis, the shikimate pathway, the cell cycle, and circadian rhythm. This study highlights the role of crop wild relatives and landraces in the diversification of long-day–adapted tetraploid potatoes and provides evidence that natural populations are a valuable source of adaptive potential.

Illumina technology: HiSeq 2000/2500 Systems, MiSeq System

Romero Navarro JA, Willcox M, Burgueño J, Romay C, et al. A study of allelic diversity underlying flowering-time adaptation in maize landraces. *Nat Genet*. 2017;49:476-80.

Landrace species are traditional varieties, which represent important genetic diversity. Genetic gains have been stagnant due to genetic linkage between a handful of useful alleles and many undesirable ones. To better understand the diversity of maize landraces, over 4000 individuals were genotyped. Over 1000 genes were identified that are associate with latitudinal and altitudinal adaptation. F-one association mapping (FOAM) analysis was used to map the genes controlling flowering across 22 environments. A large number of SNPs associated with altitude were also associated with flowering time and many altitude-associated SNPs fell within structurally variable regions. The strategy presented in this study can be used to understand the genetic basis of adaptation in important landrace species.

Illumina technology: HiSeq System

Brito LF, Clarke SM, McEwan JC, Miller SP, et al. Prediction of genomic breeding values for growth, carcass and meat quality traits in a multi-breed sheep population using a HD SNP chip. *BMC Genet*. 2017;18:1-17.

The New Zealand sheep industry relies on unique composite sheep breeds, which have high genetic diversity. Genomic selection provides an opportunity to improve carcass and meat quality traits in order to increase the rate of genetic improvement. Predicted breeding values for growth, carcass and meat quality traits were evaluated for accuracy using high-density Ovine genotyping arrays. Over 14,000 animals were genotyped and phenotypes recorded for analysis of alternative genomic relationship matrices, validation designs and genomic prediction scenarios. A mixed training population was advantageous and further analysis of breed group specific predictions will require genotyping additional individuals from each breed group to increase the size of the training population within each breed group. The level of accuracy achieved in the study support the implementation of genomic selection for improvement of New Zealand Terminal Sire breeds.

Illumina technology: OvineSNP50 BeadChip, OvineHD BeadChip

Nelore Genotyping Allows Brazilian Beef Industry to Flourish

The National Association of Breeders and Researchers (ANCP) is speeding up the evolutionary process to enhance the Nelore breed and grow the Brazilian economy.[114]

"With genomics, the ANCP is able to keep Brazilian cattle breeders at the forefront of applying new technology."

-R.B. Lôbo (ANCP)

114. *Illumina Nellore Genotyping Allows Brazilian Beef Industry to Flourish*. 2016; https://www.illumina.com/science/customer-stories/icommunity-customer-interviews-case-studies/lobo-ancp-interview-nellore-genotyping.html.

Magalhães AFB, Schenkel FS, Garcia DA, Gordo DGM, et al. Genomic selection for meat quality traits in Nelore cattle. *Meat Sci*. 2019;148:32-7.

More than 80% of cattle in Brazil are composed of the Nelore breed and its crosses with other Zebu breeds. These hardy cattle have large variation in meat quality traits, which are difficult to improve by traditional selection and breeding strategies. However, genomic selection provides an opportunity to efficiently improve meat quality traits. Approximately 5000 individuals were genotyped using arrays representing >400,000 SNPs in conjunction with phenotypic analysis of meat quality traits (i.e., tenderness, marbling, lipid percentage and meat color). The data were used to evaluate genomic prediction accuracy and bias using three different computational methods. The results demonstrate the feasibility of using genomic selection for the improvement of Nelore cattle.

Illumina technology: Bovine HD BeadChip

Figure 13. Genetic improvement of the hardy Nelore cattle breed (*B. taurus indicus*) is increasing with the implementation of genomic selection.

Available Resources

Genomic & Open-source Breeding Informatics Initiative (GOBii):[115] http://gobiiproject.org/

Germinate3:[116] https://ics.hutton.ac.uk/get-germinate/

Wheat@URGI portal:[117] https://wheat-urgi.versailles.inra.fr/

Rice Information GateWay (RIGW):[118] http://rice.hzau.edu.cn/

115. Friedmann M, Asfaw A, Anglin N, et al. Genomics-Assisted Breeding in the CGIAR Research Program on Roots, Tubers and Bananas (RTB). *Agriculture*. 2018;8:89.
116. Shaw PD, Raubach S, Hearne SJ, et al. Germinate 3: Development of a Common Platform to Support the Distribution of Experimental Data on Crop Wild Relatives. *Crop Science*. 2017;57:1259-1273.
117. Alaux M, Rogers J, Letellier T, et al. Linking the International Wheat Genome Sequencing Consortium bread wheat reference genome sequence to wheat genetic and phenomic data. *Genome Biology*. 2018;19:111.
118. Song J-M, Lei Y, Shu C-C, et al. Rice Information GateWay: A Comprehensive Bioinformatics Platform for Indica Rice Genomes. *Molecular Plant*. 2018;11:505-507

BEYOND SELECTION: USING GENOMICS FOR RESOURCE MONITORING AND MANAGEMENT

NGS has revolutionized breeding strategies for crops and livestock by allowing unprecedented access to genomic information, including non-model organisms. Beyond its application in enhanced breeding strategies, genomics can provide valuable information about the biological status of important resources like fisheries, crop and livestock health, and food safety and authenticity.

NGS is being used to identify organisms present within various environments in order to understand ecosystem diversity.[119] Species contribute DNA to their environment, which can be easily recovered and is often referred to as environmental DNA (eDNA). Species that comprise a particular environment or sample can be identified from eDNA using "DNA barcoding"—a means of differentiating species based on a unique gentic fingerprint.[120] In this way, eDNA is used to determine the repertoire of organisms present in any setting—from seawater to soil and food. This and other emerging applications of genomics are shaping best practices for resource monitoring and management related to agriculture.

Fisheries and aquaculture

Seafood is an important source of animal protein for humans, and its role in meeting future nutritional needs is growing. Dwindling natural fisheries and increasing demand have led to aquaculture becoming the fastest growing area of agricultural food production.[121] Genomics can inform the management of fisheries and aquaculture, but its application has lagged behind that of crops and livestock. Genomics can provide new insights about adaptive variaiton and enhance breeding and management strategies that promote productivity and sustainability of fisheries and aquaculture. Genomic data has been applied to understand the spatial and structural dynamics of aquatic populations, classify aquatic species, understand the genetic basis of beneifical aquaculture traits, and guard against illegal fishing and food fraud.

Reviews

Bernatchez L, Wellenreuther M, Araneda C, et al. Harnessing the Power of Genomics to Secure the Future of Seafood. *Trends Ecol Evol*. 2017;32:665-680.

Abdelrahman H, ElHady M, Alcivar-Warren A, et al. Aquaculture genomics, genetics and breeding in the United States: Current status, challenges, and priorities for future research. *BMC Genomics*. 2017;18:1-23.

McIntyre PB, Reidy Liermann CA and Revenga C Linking freshwater fishery management to global food security and biodiversity conservation. *Proceedings of the National Academy of Sciences*. 2016;113: 12880-12885.

Casey J, Jardim E and Martinsohn JTH The role of genetics in fisheries management under the E.U. common fisheries policy. *Journal of Fish Biology*. 2016;89:2755-2767.

Nielsen EE, Cariani A, Aoidh EM, et al. Gene-associated markers provide tools for tackling illegal fishing and false eco-certification. *Nature Communications*. 2012;3:851-856.

Pawlowski J, Kelly-Quinn M, Altermatt F, et al. The future of biotic indices in the ecogenomic era: Integrating (e)DNA metabarcoding in biological assessment of aquatic ecosystems. *Sci Total Environ*. 2018;637-638: 1295-1310.

119. Thomsen PF and Willerslev E Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*. 2015;183:4-18.
120. Cristescu ME and Hebert PDN Uses and Misuses of Environmental DNA in Biodiversity Science and Conservation. *Annual Review of Ecology, Evolution, and Systematics*. 2018;49:209-230.
121. *FAO The State of World Fisheries and Aquaculture 2018 - Meeting the sustainable development goals. AqTHE STATE OF THE WORLD series of the Food and Agriculture Organization of the United Nations.uaculture*. 2018;35:176; www.fao.org/3/i9540en/I9540EN.pdf.

References

Thomsen PF, Møller PR, Sigsgaard EE, Knudsen SW, et al. Environmental DNA from Seawater Samples Correlate with Trawl Catches of Subarctic, Deepwater Fishes. *PLoS One*. 2016;11.

Marine fisheries are an important natural resource that are under threat from increasing fishing and environmental pressures. Current monitoring strategies rely on expensive, invasive or unreliable methods like trawling and catch reporting. This study reports the use of environmental DNA (eDNA) and metabarcoding to monitor deep-water fisheries with comparable results to invasive trawling. Seawater was collected from depths up to nearly a kilometer near the coast of Southwest Greenland. eDNA samples were PCR amplified, sequenced and classified based on their species of origin. Important commercial fishing species in addition to the Greenland shark were the most abundant detected using the eDNA sequencing approach. eDNA abundance from deep-water fish species largely correlated with biomass data obtained from trawling with the exception of the Greenland shark, which is likely to evade capture by trawling. Quantitative assessment of marine fish using eDNA and metabarcoding remains to be proven. Yet, these results highlight the use of eDNA for qualitative assessment of marine fisheries and demonstrate its advantages over conventional monitoring methods.

Illumina technology: MiSeq System

Lien S, Koop BF, Sandve SR, Miller JR, et al. The Atlantic salmon genome provides insights into rediploidization. *Nature*. 2016;533:200-205.

Salmon are a longstanding food resource for humans and serve as an important indicator of the health of North Atlantic and Pacific costal river ecosystems. A new high-quality reference genome of Atlantic salmon (*S. salar*) reveals large genomic reorganizations that coincide with transposon-mediated repeat expansions. Comparative gene expression analyses reveal more instances of neofunctionalization rather than subfunctionalization during the salmonid-specific fourth vertebrate whole-genome duplication. Due to a high level of synteny among salmonids, S. salar reference genome information is likely to be useful for a wide variety of applications including ecological, evolutionary, conservation, and agricultural production across salmonid species.

Illumina technology: GAIIx, HiSeq 2000 System, and MiSeq System

Barth JMI, Berg PR, Jonsson PR, Bonanomi S, et al. Genome architecture enables local adaptation of Atlantic cod despite high connectivity. *Molecular Ecology*. 2017;26:4452-4466.

Marine ecosystems present an opportunity to explore local ecological adaptation despite a high level of connectivity and gene flow. Marine fish like Atlantic cod (*G. morhua*) are highly connected yet demonstrate local adaptation related to important traits. Population genomics analyses based on analysis of 12,000 SNPs in conjunction with a biophysical model of larval dispersal patterns demonstrate that despite high connectivity, genetic signatures of local adaptation remain. Enrichment of a large 5 Mb chromosomal rearrangement comprising a region containing genes important for adaptation to low salinity was discovered among discrete fjord populations. These results suggest that population-level structural variations can be influenced by local selective pressures and genome architecture is an important feature that may distinguish local ecological adaptation.

Illumina technology: Infinium II 12K SNP array

Kim OTP, Nguyen PT, Shoguchi E, Hisata, K, et al. A draft genome of the striped catfish, *Pangasianodon hypophthalmus*, for comparative analysis of genes relevant to development and a resource for aquaculture improvement. *BMC Genomics*. 2018;19.

Catfish, an important source of dietary protein, are cultivated widely throughout the world. Despite their importance as a culture stock, there is a lack of genetic information available for breeding improvement. A draft genome sequence for the striped catfish (*P. hypophthalmus*) was generated to use a genetic resource for catfish aquaculture. Analysis of mitochondrial DNA sequences confirms striped catfish phylogeny. Assembly of the nuclear genome resulted in a 700 Mb reference sequence with 568 scaffolds and N50 of >14 Mb. Roughly 28,600 gene models were annotated, which is comparable to that of other catfish and zebrafish species. Comparative analysis of catfish and zebrafish genomes revealed instances of species-specific adaptation. For example, genes required for synthesis of the sunscreen compound gadusol present in zebrafish are absent from catfish species. This new draft genome assembly provides an important resource for understanding the genetic basis of adaptation in different fish species, which can inform advanced breeding strategies in aquaculture.

Illumina technology: MiSeq System, HiSeq 2500 System

Qiu C, Han Z, Li W, Ye K, et al. A high-density genetic linkage map and QTL mapping for growth and sex of yellow drum (*Nibea albiflora*). *Scientific Reports*. 2018;8.

Yellow drum (N. albiflora) is a commercially valuable aquaculture species to East Asia. While these fish have been farmed for decades, few genetic resources are available for agronomic trait improvement. A high-density genetic linkage map containing >8,000 SNPs was constructed from sequencing a large F1 family. Using this new genetic resource, comparative analysis of gene synteny between yellow drum, medaka, and zebrafish provides insights into yellow drum genome evolution. Subsequent QTL mapping revealed putative genes involved in growth and sex-determination. This linkage map provides an important genetic resource to facilitate genomics-assisted breeding strategies based to improve this important aquaculture species.

Illumina technology: HiSeq X Ten

For Research Use Only. Not for use in diagnostic procedures.

Figure 14. Aquaculture is the fastest growing area of agricultural food production.

Disease profiling and pathogen surveillance

Infectious disease presents a major challenge to crop and livestock production and sustainability, and has direct implications for human health and food security. Whole-genome sequencing, gene expression profiling, and metagenomics have provided important insights into disease epidemiology and pathogenesis, and their application to disease profiling and pathogen surveillance in agriculture is growing. NGS technologies can provide a robust means of tracking disease progression, identifying and characterizing pathogenic organisms, and understanding host-pathogen interactions (e.g., disease resistance, adaptation and virulence).

Reviews

Gardy JL and Loman NJ Towards a genomics-informed, real-time, global pathogen surveillance system. *Nat Rev Genet*. 2018;19:9-20.

Withers S, Gongora-Castillo E, Gent D, Thomas A, Ojiambo PS and Quesada-Ocampo LM Using Next-Generation Sequencing to Develop Molecular Diagnostics for (*Pseudoperonospora cubensis*), the Cucurbit Downy Mildew Pathogen. *Phytopathology*. 2016;106:1105-1116.

Weis AM, Storey DB, Taff CC, et al. Genomic comparison of Campylobacter spp. and their potential for zoonotic transmission between birds, primates, and livestock. *Applied and Environmental Microbiology*. 2016;82:7165-7175.

Zanella R Genomic Tools and Animal Health. *Vet Sci*. 2016;3:

Mozzi A, Pontremoli C and Sironi M Genetic susceptibility to infectious diseases: Current status and future perspectives from genome-wide approaches. *Infection, Genetics and Evolution*. 2017;66:286-307.

Niederwerder MC Role of the microbiome in swine respiratory disease. *Vet Microbiol*. 2017;209:97-106.

Cortey M, Diaz I, Martin-Valls GE and Mateu E Next-generation sequencing as a tool for the study of Porcine reproductive and respiratory syndrome virus (PRRSV) macro- and micro- molecular epidemiology. *Vet Microbiol*. 2017;209:5-12.

Harding JCS, Ladinig A, Novakovic P, et al. Novel insights into host responses and reproductive pathophysiology of porcine reproductive and respiratory syndrome caused by PRRSV-2. *Vet Microbiol*. 2017;209:114-123.

References

Kamath PL, Foster JT, Drees KP, Luikart G, et al. Genomics reveals historic and contemporary transmission dynamics of a bacterial disease among wildlife and livestock. *Nature Communications*. 2016;7:1-10.

Whole-genome sequencing is an emerging approach for studying population dynamics and constructing transmission pathways for bacterial pathogens. One such pathogen, *Brucella abortus*, has a rising prevalence in Greater Yellowstone cattle and elk populations. Using whole genome sequencing, cross-species transmission and spatial diffusion dynamics were investigated. These NGS-informed analyses support a model where free-ranging elk populations are a self-sustaining brucellosis reservoir and a source for livestock infection. This study provides important information on infectious disease transmission dynamics that can inform management strategies for cattle in the region.

Illumina technology: GAIIx, MiSeq System

Menardo F, Praz CR, Wyder S, Ben-David R, et al. Hybridization of powdery mildew strains gives rise to pathogens on novel agricultural crop species. *Nature Genetics*. 2016;48:201-205.

The introduction of new crops is important for agricultural diversification, providing genetic diversity and enabling enhanced production. Despite the value of genetic diversity in agriculture, it does not guarentee sustained resistence to pathogens or disease. Moreover the forces that shape pathogen adaptation to new crop species are not well understood. The wheat—rye hybrid, triticale, was resistent to the fungal pathogen powdery mildew (*B. graminis*) until 2001. Whole genome sequencing and comparative analysis of 46 powdery mildew isolates revealed specialization for wheat, rye or triticale. Specifically, a hybrid of the two mildews adapted for wheat and rye was able to infect the hybrid plant species arising from the two hosts. These results demonstrate an agriculture-dependent mechanism of pathogen adaptation to a new crop hybrid.

Illumina technology: HiSeq System

Mahmoud M, Zeng Y, Shirali M, Yin T, et al. Genome-wide pleiotropy and shared biological pathways for resistance to bovine pathogens. *PLoS One*. 2018;13.

A major obstacle toward understanding the genetic basis of pathogen resistence has been a lack of comprehensive data linking disease resistence to host genetics. Pedigree-based studies are plagued by confounding effects from related individuals, which are likely to share a common environment. However, geneome-wide analysis of SNPs using bovine genotyping arrays and Genomic-Restricted Maximum Likelihood (G-REML) evaluation has enabled estimation of heritabilities and genetic relationships that distinguish resistence to a cohort of infectious pathogens in cattle. Significant pleiotropy between pathogen resisitence and performace was demonstrated in cows and calves with the identification of genes in the B-lymphocyte pathway as important contributors to resistence for all pathogens evaluated. This study provides a framework for identifying the genetic basis of pathogen resistance with the aim of using this information to breed more productive cattle.

Illumina technology: Bovine 50K SNP-BeadChip V2, Illumina Bovine Eurogenomics 10K low-density chip

Food safety and authenticity

NGS methods like targeted amplicon and whole-genome sequencing are emerging as the next generation of food safety and authenticity testing technology. NGS-based DNA barcoding allows food producers to identify pathogens, allergens, and ingredients based on detection of a unique DNA signature—or "barcode"—from a mixed DNA sample obtained from food products.

Reviews

Deng X, den Bakker HC and Hendriksen RS Genomic Epidemiology: Whole-Genome-Sequencing–Powered Surveillance and Outbreak Investigation of Foodborne Bacterial Pathogens. *Annual Review of Food Science and Technology*. 2016;7:353-374.

Taboada EN, Graham MR, Carrico JA and Van Domselaar G Food Safety in the Age of Next Generation Sequencing, Bioinformatics, and Open Data Access. *Front Microbiol*. 2017;8:909.

Mishra P, Kumar A, Nagireddy A, et al. DNA barcoding: an efficient tool to overcome authentication challenges in the herbal market. *Plant Biotechnol J*. 2016;14:8-21.

Jagadeesan B, Gerner-Smidt P, Allard MW, et al. The use of next generation sequencing for improving food safety: Translation into practice. *Food Microbiol*. 2019;79:96-115.

Whole-Genome Sequencing Traces Origin of *E. coli* Outbreak in Romaine Lettuce

An outbreak of *E. coli* infections linked to romaine lettuce was traced back to a contaminated reservoir using whole- genome sequencing.[122] The work, conducted by the U.S. Centers for Disease Control and Prevention and the U.S. Food and Drug Administration, is part of a growing effort to use whole-genome sequencing for foodborne disease surveillance.[123]

"Whole-genome sequencing (WGS) is beginning to replace pulsed-gel electrophoresis (PFGE) for subtyping of foodborne pathogens from stools and other specimens for outbreak surveillance." -Carleton 2016

Figure 15. Whole-genome sequencing traced an outbreak of *E. coli* in romaine lettuce to its source.

122. *FDA Investigation Summary: Factors Potentially Contributing to the Contamination of Romaine Lettuce Implicated in the Fall 2018 Multi-State Outbreak of E. coli* O157:H7. 2019; https://www.fda.gov/Food/RecallsOutbreaksEmergencies/Outbreaks/ucm631243.htm
123. Carleton H and Gerner-Smidt P Whole-Genome Sequencing Is Taking over Foodborne Disease Surveillance: Public health microbiology is undergoing its biggest change in a generation, replacing traditional methods with whole-genome sequencing. *Microbe Magazine*. 2016;11:311-317.
124. Allard MW, Strain E, Melka D, et al. Practical value of food pathogen traceability through building a whole-genome sequencing network and database. *Journal of Clinical Microbiology*. 2016;54:1975-1983.

References

Dobrovolny S, Blaschitz M, Weinmaier T, Pechatschek J, et al. Development of a DNA metabarcoding method for the identification of fifteen mammalian and six poultry species in food. *Food Chemistry*. 2019;272:354-361.

Food adulteration, while illegal, remains a problem throughout the world. Food adulteration can include substitution of food ingredients, dilution of ingredients, unauthorized food processing steps and/or declaration of the incorrect geographical origin. Meat products in particular may suffer from adulteration with inferior quality products. This work describes and establishes a DNA metabarcoding approach to identify 15 mammalian and 5 poultry species by their 16S mitochondrial rRNA sequences. The species of interest was identified and differentiated from other related species 99.9% of the time. The ability to multiplex and parallelize DNA metabarcoding has the potential to enable routine application for food authenticity testing.

Illumina technology: MiSeq

Silletti S, Morello L, Gavazzi F, Gianì S, et al. Untargeted DNA-based methods for the authentication of wheat species and related cereals in food products. *Food Chemistry*. 2019;271: 410-418.

Food commodities such as pasta and breads, are often made using mixed flours that contain a variety of cereal grains. Traditional analytical methods to determine the composition of plant species in these food products include targeted qPCR analysis, which does not allow the unbiased detection of organisms. This qPCR approach can also be confounded by amplification of nearly identical DNA sequences from closely related species. However, a DNA fingerprinting approach using tubulin-based polymorphism (TBP) or TBP *light* has been optimized for the authentication of commercial food products containing wheat, farro, and other cereal grain species. The TBP assay is sensitive to detect 0.5-1% w/w in binary food mixtures of durum wheat in einkorn or emmer flour and was sufficient to authenticate test food composition and identify adulterers.

Illumina technology: MiSeq

Available Resources

Genome Trakr Network:[124] https://www.fda.gov/Food/FoodScienceResearch/WholeGenomeSequencingProgramWGS/ucm363134.htm

PulseNet:[122] https://www.cdc.gov/pulsenet/index.html

Microbial metagenomics in agriculture

Genomics technologies like NGS enable deeper exploration of the microbial communities that are ubiquitous in our environment. Metagenomic shotgun sequencing is now routinely used to classify and characterize microbial communties directly from their environment. Technological advances coupled with a growing interest in the organisms that colonize plants and animals has spurred initiatives such as the Earth Microbiome Project.[125] Plant and animal-associated microbiomes have coevolved with their hosts to form complex symbiotic relationships. Although a concerted effort to characize the diverse taxa and functions of microbiomes is underway, the application of microbiome manipulation and management strategies to agriculture will require large-scale and integrated metagenomics frameworks.[126,127]

Plants

Metagenomic sequencing of plant-associated microbiomes reveals complex community structures with different taxa fulfilling unique roles. Plant microbiomes impact host nutrition, tissue development, and disease. To date, the best studied micobial communities colonize plants from the soil to form the rhizosphere (on root surfaces) and endosphere (inside roots). For example, nitrogen-fixing bacteria associated with legume root nodules have been extensively characterized (Figure 16). Plant hosts provide an important source of carbon for associated microbes, and in turn, these communities support the productivity of their host (e.g., by providing a source of nitrogen to support growth).

There is a growing appreciation for the role microbiomes play in plant yield, abiotic stress tolerance, and disease resistance, yet the underlying mechanisms that determine plant-microbial interactions remain largely unexplored. Specifically, understanding the relationship between microbiome, host genetic, and environmental interactions will enable development of management strategies that meet productivity and sustainability goals for agriculture.

Reviews

Busby PE, Soman C, Wagner MR, et al. Research priorities for harnessing plant microbiomes in sustainable agriculture. *PLoS Biol*. 2017;15:e2001793.

Van der Heijden MGA and Hartmann M Networking in the Plant Microbiome. *PLOS Biology*. 2016;14:e1002378.

Pieterse CMJ, de Jonge R and Berendsen RL The Soil-Borne Supremacy. T*rends Plant Sci*. 2016;21:171-173.

Bakker P, Pieterse CMJ, de Jonge R and Berendsen RL The Soil-Borne Legacy. *Cell*. 2018;172:1178-1180.

Kroll S, Agler MT and Kemen E Genomic dissection of host-microbe and microbe-microbe interactions for advanced plant breeding. *Curr Opin Plant Biol*. 2017;36:71-78.

125. Thompson LR, Sanders JG, McDonald D, et al. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature*. 2017;551:457.

126. Waldor MK, Tyson G, Borenstein E, et al. Where next for microbiome research? *PLoS* Biol. 2015;13:e1002050.

127. Ikeda-Ohtsubo W, Brugman S, Warden CH, Rebel JMJ, Folkerts G and Pieterse CMJ How Can We Define "Optimal Microbiota?": A Comparative Review of Structure and Functions of Microbiota of Animals, Fish, and Plants in Agriculture. *Front Nutr*. 2018;5:90

For Research Use Only. Not for use in diagnostic procedures.

Figure 16. Microbial communities that colonize plant roots provide important nutrients. Rhizobia within root nodules of legumes utilize carbon present in root exudates and in turn fix atmospheric nitrogen, which benefits their host.

References

Castrillo G, Teixeira PJ, Paredes SH, Law TF, et al. Root microbiota drive direct integration of phosphate stress and immunity. Nature. 2017;543:513-518.

Plants occupy diverse habitats that contain unique soils and microbiota. These microbial communities colonize the exterior and interior of plant organs like roots and can change in response to environmental conditions such as the nutrient content of soils. Plant-associated microbial communties are also shaped by symbiotic and competitive interactions between both their plant hosts and other microbial taxa present. These complex interactions between host immune systems, environment and microbiota ultimately impact community structure and assembly, but are not well-understood. This study describes a genetic network controlling phosphate stress response in both stress and non-stress conditions, and identifies a molecular mechanism responsible for balancing nutrition versus defense in the presence of a synthetic microbial community. The results suggest that plants prioritize nutrition over defense, which has important implications for harnessing microbiomes to enhance plant performance.

Illumina technology: MiSeq System, HiSeq 2500 System

Levy A, et al. Genomic features of bacterial adaptation to plants. *Nature Genetics*. 2018;50:138–150.

Microbes have adaped to living in and around plants for millions of years. The genes that shape plant-microbal interactions and mutual adaptation are not well understood. Whole genome sequencing of bactrial isolates from Brassicaceae, poplar and maize reveal thousands of plant-associated gene clusters. These bacterial genomes tend to encode more genes involved in carohydrate metabolism—likely important for adaptation to plant environments. Their genomes also tend to have fewer mobile elements. Genes involved in plant colonization or microbe-microbe interactions were further validated using publically available metagenome samples. This work also led to the discovery of protein domains in plant microbiomes that mimic those found in host innate immune system proteins. These proteins may alow plant-associted microbes to augment host immunity. Plant-associted function in microbes are consistent across many taxa. These insights into microbial adaptation pants can inform microbiome engieering strategies for agriculture.

Illumina technology: HiSeq 2000/2500 Systems

Kwak M, et al. Rhizosphere microbiome structure alters to enable wilt resistance in tomato. *Nature Biotechnology*. 2018;36:1100–1109.

The innate immune system of plants detects microbe-associated molecular patterns and shapes the plant's response to microbial colonization. Rhizoshpere communities are comprised of large numbers of diverse taxa, but the structure and function of these microbiomes in confering disease resistence is unclear. Tomato plants provide a tractable model to study bacterial wilt, a disease caused by the soil-borne bacteria *R. solanaceae*. These bacteria enter roots invading xylem, blocking water transport and ultimately leading to wilt. Analysis of the microbiomes of two tomato cultivars using metagenomic shotgun sequencing in conjunction with microbiome transplantation studies reveals a rhizohpere microbiome strucure that confers wilt resistance in tomato. These results indicate a role for native microbiota in protecting plants from pathogenic microorganisms, which may enable the development of probiotics to protect crops from disease.

Illumina techology: HiSeq 2000 System

Fitzpatrick CR, et al. Assembly and ecological function of the root microbiome across angiosperm plant species. *Proceedings of the National Academy of Sciences*. 2018;115: E1157–E1165.

Root microbial communities have shaped the evolution of land plants and are key members of the plant ecosystem. The functions that drive mutual adaptation and confer beneticial or detrimental interactions between hosts and the species that colonize them remain unclear. Analysis of root microbiomes associated with 30 different angiosperm species spanning 140 million years of evolution reveals comparable diversity among the microbial communities that inhabit them. Microbial communities affect host adaptation to abiotic stress and in turn influence competitive interactions between plants occupying the same ecosystems. These results highlight the factors that influence root microbiome variation and their importatnce to abiotic and biotic stress adaptation.

Illumina technology: MiSeq System

Available Resources

The Earth Microbiome Project:[125] http://www.earthmicrobiome.org/

Animals

Like plants, animals live in intimate association with microbial communities, which colonize predominantly the surfaces of tissues such as the respiratory tract and gut. Animal microbiomes function in similar ways to those of plants, influencing host nutrition, growth, and health. Metagenomics of animal microbiomes has led to a better understanding of microbial composition and function, especially for livestock such as cattle. Microbiome information in conjunction with data on host nutrition and physiology, are enabling development of new intervention strategies to increase productivity and sustainability in agriculture.

Reviews

McFall-Ngai M, Hadfield MG, Bosch TC, et al. Animals in a bacterial world, a new imperative for the life sciences. *Proc Natl Acad Sci U S A*. 2013;110:3229-3236.

Malmuthuge N and Guan LL Gut microbiome and omics: a new definition to ruminant production and health. *Animal Frontiers*. 2016;6:8-12.

Alexander TW and Plaizier JC The importance of microbiota in ruminant production. *Animal Frontiers*. 2016;6:4-7.

Matthews C, Crispie F, Lewis E, Reid M, O'Toole PW and Cotter PD The rumen microbiome: a crucial consideration when optimising milk and meat production and nitrogen utilisation efficiency. *Gut Microbes*. 2018;00:1-18.

Zeineldin M, Barakat R, Elolimy A, Salem AZM, Elghandour MMY and Monroy JC Synergetic action between the rumen microbiota and bovine health. *Microb Pathog*. 2018;124:106-115.

References

Roehe R, Dewhurst RJ, Duthie C, Rooke JA, et al. Bovine Host Genetic Variation Influences Rumen Microbial Methane Production with Best Selection Criterion for Low Methane Emitting and Efficiently Feed Converting Hosts Based on Metagenomic Gene Abundance. *PLoS Genetics*. 2016;12:2.

Methane—a potent greehouse gas—is largely produced by methanogenic archea in ruminants like cattle. Yet, the host genetic features that dictate microbial methane production are unknown and genetic selection strategies for mitigating rumen methane production remain undeveloped. Sire progeny group differences were used to determine the host genetic influence on rumen microbial methane production in conjuntion with varying diets. Metagenomic analysis of the rumen microbiome was conducted to determine the links between microbial genes, methane emmisions and feed conversion efficiency. Sire progeny groups showed significant differences in their methane emissions and ranking of animals based on methane emissions was consistent with ranking based on relative archeal abundance regardless of diet. These results suggest that archeal abundance in the rumen microbiome is considerably influenced by host genetics, and animals could be subject to genetic selection for methane production without measuring methane directly. Microbial genes involed in methane production and feed conversion efficiency were identified and explained a majority of the variance seen between host microbiomes. Methanogenesis genes were associated with methane production while host micrbiome crosstalk genes were associated with feed conversion efficency. These results suggest that host genetic influences play a major role in microbiome composition and function. Given the link between host genetics, rumen microbial gene expression and methane production, genetic selection strategies could be implemented to alleviate methane production in ruminants.

Illumina technology: HiSeq 2500 System

For Research Use Only. Not for use in diagnostic procedures.

Mitra N, Cernicchiaro N, Torres S, Li F, et al. Metagenomic characterization of the virome associated with bovine respiratory disease in feedlot cattle identified novel viruses and suggests an etiologic role for influenza D virus. *Journal of General Virology*. 2016;97: 1771–1784.

Bovine respiratory diease is a major problem affecting productivity in cattle. Bovine respiratory disease etiology is complex—influenced by microbial, host and environmental factors. Viral metagenmoic sequencing of swabs from feedlot cattle with and without disease revealed an association between influenza D virus and affeted cattle. Data gathered from this study highlights the complexity of the virome associated with bovine respirat disease and the potnetial contribution of particular viruses to disease pathogenesis.

Illumina technology: MiSeq System

A Catalog of Reference Genomes from the Rumen Microbiome

The Huntagate1000 project, named after prominent rumen microbiologist, Bob Hungate, aims to sequence 1000 rumen microbial genomes in an effort to establish a community resource for rumen metagenomics. This reference genome information will be valuable for the interpretaion of rumen metagenomics datasets and ultimately further our understanding of rumen biology.[128]

Seshadri, R. et al. Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. *Nature Biotechnology*. 2018;36:359-367.

The rumen functions as the primary site of microbial fermentation in the digestive tract of animals like cattle. Bacteria present in this comparment fermet otherwise indigestible plant cellulosic material into accessible nutrients. Thus, rumen microbiota are useful to study greenhouse gas production in ruminants and develop better biofuels for lignocellulose. The Huntgate 1000 project aims to provide a reference collection of genomes from all microorganisms present in the rumen of livestock. Metagenomic sequencing of bacterial and archeal organisms from rumen environments shows an enrichment of species capable of de novo synthesis of vitamin B12 and indicates ongoing evolution and vertical inheritance of the rumen microbiome. This work presents a valuable resource for understanding how carbon is efficiently utilized by rumen microbiota and how this process impacts ruminant productivity and sustainability.

Illumina technology: HiSeq System, NextSeq System, MiSeq System

"This vital reference information supports international efforts to develop methane mitigation and rumen adaptation technologies, as well as to further genome-enabled research aimed at understanding rumen function, feed conversion efficiency, methanogenesis and plant cell wall degradation in order to find a balance between food production and greenhouse gas emissions." -RMG Network

Figure 17. Understanding the rumen microbiome is integral to efforts to improve cattle productivity and sustainability.

Available Resources

Rumen Microbial Genomics Network:[128] http://www.rmgnetwork.org/

128. *RMG Network 2016*; http://www.rmgnetwork.org/

THE FUTURE OF AGRIGENOMICS

NGS technologies have enabled the unprecedeted collection of genomic information on thousands of previously intractable organisms. Falling costs and improved tools have led to widespread adoption of NGS to interrogate plants, animals and their associated microbiota with higher precision and breadth than previously possible. Access to high-quality genomic information has led to the development of cost-efficient genotyping technologies and identification of the genetic basis of valuable agronomic traits. Reference genome sequences also allow application of gene editing technologies, which can rapidly introduce desirable traits into plants and animals. This wealth of genomic information is guiding advanced breeding strategies that incorporate the *holobiont*—the collection of species forming an ecosystem.[129] The future of agrigenomics is the application of genomic (and metagenomic) engineering to enhance crops and livestock. These advanced breeding strategies are vital to meet mounting production and sustainability requirements for agriculture.

129. Sánchez-Cañizares C, Jorrín B, Poole PS and Tkacz A Understanding the holobiont: the interdependence of plants and their microbiome. *Current Opinion in Microbiology*. 2017;38:188-196

Gene editing

Crop and livestock improvement is often a slow and laborious process using traditional approaches such as genetic recombination or random mutagenesis, and cannot keep up with increasing demands for sustainable and nutritious food. Genome editing technologies like CRISPR/Cas (clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated protein [CRISPR/Cas]) and TALENs (transcription activator-like effector nucleases [TALENs]) now allow targeted modification of genome sequences to generate novel and/or useful variation Gene editing piplines rely on NGS technologies for producing a reference genome, identifying targets for genome editing using functional genomics, and assessing the specificity of genome edited lines.

Reviews

Scheben A, Wolter F, Batley J, Puchta H and Edwards D Towards CRISPR/Cas crops - bringing together genomics and genome editing. *New Phytologist*. 2017;216:682-698.

Van Eenennaam AL Genetic modification of food animals. *Curr Opin Biotechnol*. 2017;44:27-34.

Ricroch A, Clairand P and Harwood3 W Use of CRISPR systems in plant genome editing: toward new opportunities in agriculture. *Emerginig Topics in Life Sciences*. 2017;1:169-182.

Voytas DF and Gao C Precision genome engineering and agriculture: opportunities and regulatory challenges. *PLoS Biol*. 2014;12:e1001877.

References

Carlson DF, Lancto CA, Zang B, et al. Production of hornless dairy cattle from genome-edited cell lines. *Nature Biotechnology*. 2016;34:479-481.

Physical dehoring of dairy cattle is performed to reduce the chance of animal and handler injury, yet the procedure is costly and painful for the animals. Beef cattle breeds have a higher incidence of the *POLLED* locus, which results in naturally hornless (i.e., polled) cattle. There is a growing demad for naturally polled dairy cattle that do not require dehorning. While there are polled dairy cattle bulls available they have much lower EBVs, which would require multiple decades to improve using traditional breeding strategies. Using genomics, researchers discovered alleles that have arisen 500-100 years ago and confer the polled phentype. Using transcription activator-like effector nucleases (TALEN) gene editing technology researchers introduced the *POLLED* allele into the genomes of bovine embryonic fibroblasts. After characterization of the gene-edited cell lines, three were selected for somatic cell nuclear transfer and implantation into receptive cows. Five live offspring were generated all of which showed the polled phenotype.

Illumina technology: HiSeq 2500 System

Soyk S, et al. Bypassing Negative Epistasis on Yield in Tomato Imposed by a Domestication Gene. *Cell*. 2017;169:1142–1155.e12.

Inflorescence architecture—the arrangement of flowers on a plant—is generally a highly selected trait in domesticated crops. Yet domesticated tomato inflorescence structures more closely resemble their wild counterparts. This is largely due to a link between excessive branching and low fertiliy. Plants carrying two mutations—one that enhances fruit size and another producing stems that facilitate harvest—result in undesirable inforescence architecture and sterility due to an epistatic mechanism. Gene editing using CRISPR/Cas enabled both investigation and amelioration of an epistatic mechanism linking desireable and undesirable agronomic traits in domesticaed tomatoes. This work highlights the utility of gene editiing to understand and exploit negative epistasis, which can significantly affect crop productivity.

Illumina technology: HiSeq 2500 System

Advanced breeding strategies

In addition to gene editing, microbiome interventions are a promsing approach to enhance crop and livestock productivity, health, and sustainability. Genomic and metagenomic engineering are the next-generation breeding strategies to accelerate genetic gains in crop and livestock species.

Reviews

Gopal M and Gupta A Microbiome Selection Could Spur Next-Generation Plant Breeding Strategies. *Front Microbiol*. 2016;7:1971.

Kang YJ, Lee T, Lee J, et al. Translational genomics for plant breeding with the genome sequence explosion. *Plant Biotechnology Journal*. 2016;14:1057-1069.

Wei Z and Jousset A Plant Breeding Goes Microbial. *Trends Plant Sci*. 2017;22:555-558.

Kroll S, Agler MT and Kemen E Genomic dissection of host-microbe and microbe-microbe interactions for advanced plant breeding. *Curr Opin Plant Biol*. 2017;36:71-78.

Kong Z, Hart M and Liu H Paving the Way From the Lab to the Field: Using Synthetic Microbial Consortia to Produce High-Quality Crops. *Front Plant Sci*. 2018;9:1467.

Hu H, Scheben A and Edwards D Advances in Integrating Genomics and Bioinformatics in the Plant Breeding Pipeline. *Agriculture*. 2018;8:75.

Li H, Rasheed A, Hickey LT and He Z Fast-Forwarding Genetic Gain. *Trends Plant Sci*. 2018;23:184-186.

References

Mitter B, et al. A New Approach to Modify Plant Microbiomes and Traits by Introducing Beneficial Bacteria at Flowering into Progeny Seeds. *Frontiers in Microbiology*. 2017;8:11.

The seed microbiome is thought to be inherited between generations of plants, influencing germination and productivity. Thus, methods that optimize seed microbiomes can be useful for crop breeding strategies to enhance food, feed and fiber production. A new approach identifies favorable traits mitigated by seed microbiomes and enables augmentation of seed microbiomes in elite crop species. Introduction of the endophyte P. phytofirmans to the flowers of parental plants led to the inclusion of this organism in progeny seed microbiomes, which confirms a model of vertical inheritance. Introduction of P. phytofirmans to seeds of monocot and dicot plant species led to changes in seed microbiome composition that utimately affected growth traits in wheat. Ths work illustrates the potential for metagenome engineering to influence agronomic trains in crop species.

Illumina technology: MiSeq System

Ueta R, Abe C, Watanabe T, Sugano S, et al. Rapid breeding of parthenocarpic tomato plants using CRISPR/Cas9. *Scientific Reports*. 2017;7:507.

Parthenocarpy—the ability to produce fruit without seeds or fertilization—is a valuable agronomic trait. Seedless fruit are beneficial for crop production in fluctuating environments and have added end-use value in food production. This study demonstrates an advanced breeding strategy using CRISPR/Cas to rapidly generate parthenocarpic tomato plants. An optimized CRISPR/Cas system was used to introduce somatic mutations into the SlIAA9 gene. A mutation rate of 100% was achieved in the first generation (T0) and deep genome sequencing did not reveal notable off-target effects. Regenerated mutant tomatoes had phenotypic changes consistent with parthenocarpy—changes in leaf shape and seedless fruit. The segregated next generation (T1) of plants maintained the parthenocarpic phenotype of the first generation of mutated plants. This work establishes the use of CRISPR/Cas gene editing to rapidly and precisely develop parthenocarpic crops.

Illumina technology: MiSeq System

Figure 18. CRISPR/Cas gene editing technology has been successfully applied to engineer beneficial agronomic traits in tomatoes.

For Research Use Only. Not for use in diagnostic procedures.

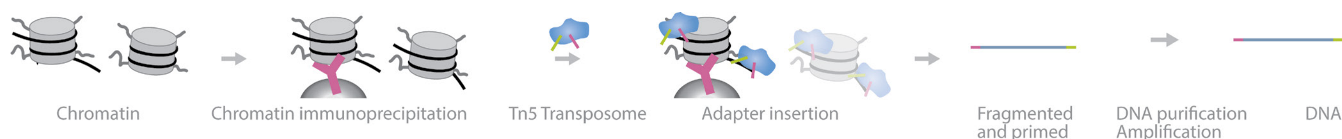
Genome assembly and epigenetics

ChIPmentation: Chromatin Immunoprecipitation with Sequencing Library Preparation by Tn5 Transposase

ChIPmentation combines ChIP with sequencing library preparation by Tn5 transposase (tagmentation).[130] Both ATAC-Seq and ChIPmentation can be combined usefully in the same experiment to assay open chromatin and protein binding, respectively.[131] This combination of methods takes advantage of the flexibility and efficiency provided by using Tn5 transposase in library preparation.

130. Schmidl C, Rendeiro AF, Sheffield NC and Bock C ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nat Methods*. 2015;12:963-965.
131. Chaitankar V, Karakulah G, Ratnapriya R, Giuste FO, Brooks MJ and Swaroop A Next generation sequencing technology and genomewide data analysis: Perspectives for retinal research. *Prog Retin Eye Res*. 2016;55:1-31.

Tagmentation is performed directly on bead-bound, immunoprecipitated chromatin, followed by standard library preparation methods.



A schematic overview of ChIPmentation.

Table 4. Advantages and Disadvantages of ChIPmentation

Advantages	Disadvantages
• Can generate accurate profiles from as little as 10,000 cells • Simple, one-step reaction	• None reported

References

Rendeiro AF, Schmidl C, Strefford JC, et al. Chromatin accessibility maps of chronic lymphocytic leukaemia identify subtype-specific epigenome signatures and transcription regulatory networks. *Nat Commun*. 2016;7:11938.

The authors studied epigenetic deregulation in chronic lymphocytic leukemia (CLL). Chromatin profiles created by ChIPmentation and RNA-Seq accurately predicted CLL. The authors also examined the mutation status of IGHV genes, a clinical biomarker for CLL. Gene regulatory networks inferred for IGHV-mutated vs IGHV-unmutated samples identified characteristic differences between less aggressive and more aggressive CLL subtypes, respectively.

Illumina technology: HiSeq 3000/4000 Systems

Associated Kits

TruSeq ChIP Library Prep Kit

TruSeq Nano DNA Library Prep Kit

TruSeq DNA Sample Preparation Kit

TruSeq DNA PCR-Free Library Prep Kit

Nextera® DNA Library Prep Kit

Nextera XT DNA Library Prep Kit

ChIA-PET: Chromatin Interaction Analysis by Paired-End Tag Sequencing

ChIA-PET features an immunoprecipitation step to map long-range DNA interactions, similar to Hi-C.[132,133] In this method, DNA-protein complexes are crosslinked and fragmented. Specific antibodies are used to immunoprecipitate proteins of interest. Two sets of linkers, with unique barcodes, are ligated to the ends of the DNA fragments in separate aliquots, which then self-ligate based on proximity. The DNA aliquots are precipitated, digested with restriction enzymes, and sequenced. Deep sequencing provides base-pair resolution of the ligated fragments. Hi-C and ChIA-PET currently provide the best balance of resolution and reasonable coverage in the human genome to map long-range interactions.[134]

A modified protocol, called advanced or long-read ChIA-PET, has been published by Tang et al.[135] This method replaces the 2 separate ligation reactions with 2 half linkers and a single biotinylated linker ligation. Next, the de-crosslinked, purified DNA is fragmented, and adapters are ligated using Tn5 transposase in a single step. Finally, the DNA is PCR-amplified and sequenced.[136]

132. Li G, Fullwood MJ, Xu H, et al. ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. *Genome Biol*. 2010;11:R22.

133. Fullwood MJ, Liu MH, Pan YF, et al. An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature*. 2009;462:58-64.

134. Dekker J, Marti-Renom MA and Mirny LA Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat Rev Genet*. 2013;14:390-403.

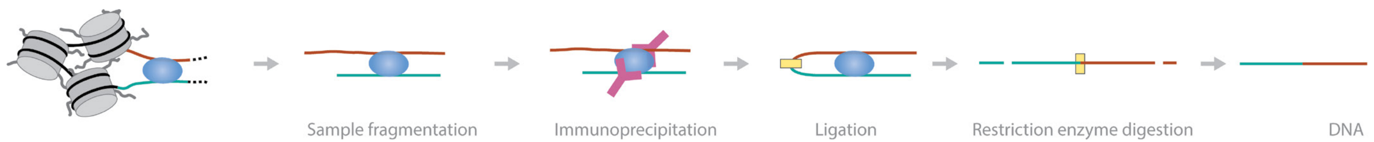
135. Tang Z, Luo OJ, Li X, et al. CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell*. 2015;163:1611-1627.

136. Sati S and Cavalli G Chromosome conformation capture technologies and their impact in understanding genome function. *Chromosoma*. 2016;

137. Sajan SA and Hawkins RD Methods for identifying higher-order chromatin structure. *Annu Rev Genomics Hum Genet*. 2012;13:59-82.

138. Consortium EP A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol*. 2011;9:e1001046.

139. Mumbach MR, Rubin AJ, Flynn RA, et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat Methods*. 2016;13:919-922.



A schematic overview of ChIA-PET.

Table 5. Advantages and Disadvantages of ChIPmentation

Advantages	Disadvantages
• Suitable for detecting a large number of both long-range and short-range chromatin interactions globally[137]	• Requires a large amount of starting material generally at least 100 million cells[139]
• Studies the interactions made by specific proteins or protein complexes	• Nonspecific antibodies can pull down unwanted protein complexes and contaminate the pool
• Public ChIA-PET datasets are available through the ENCODE Project[138]	• Linkers can self-ligate, generating ambiguity about true DNA interactions
• Removes background generated during traditional ChIP assays	• Limited sensitivity; may detect as little as 10% of interactions
• Immunoprecipitation step reduces data complexity	• Requires a large amount of starting material generally at least 100 million cells

Associated Kits

TruSeq ChIP Library Prep Kit

TruSeq Nano DNA Library Prep Kit

TruSeq DNA Sample Prep Kit

TruSeq DNA PCR-Free Library Prep Kit

Nextera DNA Library Prep Kit

Nextera XT DNA Library Prep Kit

Reviews

Sati S and Cavalli G Chromosome conformation capture technologies and their impact in understanding genome function. *Chromosoma*. 2016;.

Chang H, Liu Y, Xue M, et al. Synergistic action of master transcription factors controls epithelial-to-mesenchymal transition. *Nucleic Acids Res*. 2016;44:2514-2527.

Darabi H, Beesley J, Droit A, et al. Fine scale mapping of the 17q22 breast cancer locus using dense SNPs, genotyped within the Collaborative Oncological Gene-Environment Study (COGs). *Scientific reports*. 2016;6:32512.

Fujimoto A, Furuta M, Totoki Y, et al. Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. *Nat Genet*. 2016;48:500-509.

Tewhey R, Kotliar D, Park DS, et al. Direct Identification of Hundreds of Expression-Modulating Variants using a Multiplexed Reporter Assay. *Cell*. 2016;165:1519-1529

References

Ricano-Ponce I, Zhernakova DV, Deelen P, et al. Refined mapping of autoimmune disease associated genetic variants with gene expression suggests an important role for non-coding RNAs. *Journal of autoimmunity*. 2016;68:62-74.

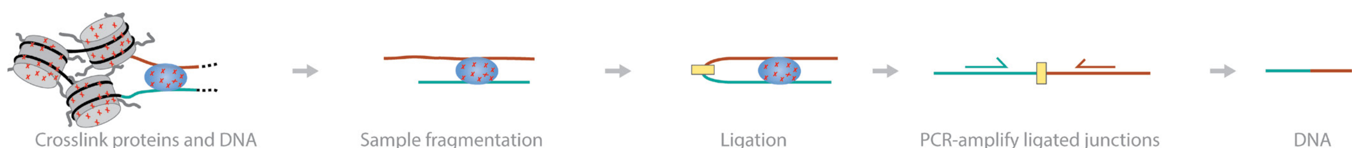
The authors performed a systematic analysis of transcriptomic data from 629 blood samples and linked 460 SNPs that are associated with 14 autoimmune diseases. The SNPs were linked to functional regulatory elements, which suggests a model where autoimmune disease genes are regulated by a network of chromatin-looping/noncoding RNA interactions. Data from the ChIA-PET assay in B lymphoblastoid cells indicated that multigene SNPs were more often (70%) involved in looping interactions in B cells than single-gene SNPs (55%).

Illumina Technology: HiSeq 2000 System

3-C, Capture-C, and Hi-C: Chromatin Conformation Capture Sequencing

3C-Seq,[140] Capture-C, and Hi-C[141] comprise a family of methods for analyzing chromatin interactions. Capture-C adds an additional pull-down of the biotinylated fragments with magnetic beads to the 3C method. A new refinement of the Capture-C method (NG Capture-C) is available.[142] The Hi-C approach extends 3C-Seq to map chromatin contacts genome-wide, and it has also been applied to studying in situ chromatin interactions.[143,144]

In this method, DNA-protein complexes are crosslinked with formaldehyde. The sample is fragmented, and the DNA is extracted, ligated, and digested with restriction enzymes. The resulting DNA fragments are PCR-amplified and sequenced. Deep sequencing provides base-pair resolution of the ligated fragments.



A schematic overview of 3C-Seq.

140. Duan Z, Andronescu M, Schutz K, et al. A genome-wide 3C-method for characterizing the three-dimensional architectures of genomes. *Methods*. 2012;58:277-288.

141. Lieberman-Aiden E, van Berkum NL, Williams L, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 2009;326:289-293.

142. Davies JO, Telenius JM, McGowan SJ, et al. Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nat Methods*. 2016;13:74-80.

143. Sati S and Cavalli G Chromosome conformation capture technologies and their impact in understanding genome function. *Chromosoma*. 2016;

144. Rao SS, Huntley MH, Durand NC, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014;159:1665-1680.

145. Bourgo RJ, Singhal H and Greene GL Capture of associated targets on chromatin links long-distance chromatin looping to transcriptional coordination. *Nat Commun*. 2016;7:12893.

Table 6. Advantages and Disadvantages of 3C-Seq

Advantages	Disadvantages
• Allows detection of long-range DNA interactions • High-throughput method	• Detection may result from random chromosomal collisions • Less than 1% of DNA fragments actually yield ligation products[145] • Due to multiple steps, the method requires large amounts of starting material

Reviews

Sati S and Cavalli G Chromosome conformation capture technologies and their impact in understanding genome function. *Chromosoma*. 2016;

Turaev D and Rattei T High definition for systems biology of microbial communities: metagenomics gets genome-centric and strain-resolved. *Curr Opin Biotechnol*. 2016;39:174-181.

Acemel RD, Tena JJ, Irastorza-Azcarate I, et al. A single three-dimensional chromatin compartment in amphioxus indicates a stepwise evolution of vertebrate Hox bimodal regulation. *Nat Genet*. 2016;48:336-341.

Bigot P, Colli LM, Machiela MJ, et al. Functional characterization of the 12p12.1 renal cancer-susceptibility locus implicates BHLHE41. *Nat Commun*. 2016;7:12098.

Capurso D, Bengtsson H and Segal MR Discovering hotspots in functional genomic data superposed on 3D chromatin configuration reconstructions. *Nucleic Acids Res*. 2016;44:2028-2035.

Gunnell A, Webb HM, Wood CD, et al. RUNX super-enhancer control through the Notch pathway by Epstein-Barr virus transcription factors regulates B cell growth. *Nucleic Acids Res*. 2016;44:4636-4650.

Kim KD, Tanizawa H, Iwasaki O and Noma K Transcription factors mediate condensin recruitment and global chromosomal organization in fission yeast. *Nat Genet*. 2016;48:1242-1252.

Petryk N, Kahli M, d'Aubenton-Carafa Y, et al. Replication landscape of the human genome. Nat Commun. 2016;7:10208.

Pichugina T, Sugawara T, Kaykov A, et al. A diffusion model for the coordination of DNA replication in Schizosaccharomyces pombe. *Scientific reports*. 2016;6:18757.

Putnam NH, O'Connell BL, Stites JC, et al. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res*. 2016;26:342-350.

Rocha PP, Raviram R, Fu Y, et al. A Damage-Independent Role for 53BP1 that Impacts Break Order and Igh Architecture during Class Switch Recombination. *Cell Rep*. 2016;16:48-55.

Tjong H, Li W, Kalhor R, et al. Population-based 3D genome structure analysis reveals driving forces in spatial genome organization. *Proc Natl Acad Sci USA*. 2016;113:E1663-1672.

Ulianov SV, Khrameeva EE, Gavrilov AA, et al. Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome Res*. 2016;26:70-84.

References

Criscione SW, De Cecco M, Siranosian B, et al. Reorganization of chromosome architecture in replicative cellular senescence. *Sci Adv*. 2016;2:e1500882.

The authors used Hi-C, fluorescence in situ hybridization (FISH), and in silico modeling methods to characterize the 3D architecture of interphase chromosomes in proliferating, quiescent, and senescent cells. Direct measurements of distances between genetic loci, chromosome volumes, and chromatin accessibility suggested that the Hi-C interaction changes were caused by a significant reduction of the volumes occupied by individual chromosome arms. In contrast, centromeres opposed this overall compaction trend and increased in volume.

Illumina Technology: HiSeq 2500 System

Darrow EM, Huntley MH, Dudchenko O, et al. Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. *Proc Natl Acad Sci USA*. 2016;113:E4504-4512.

During interphase, the inactive X chromosome (Xi) is largely silent transcriptionally and adopts an unusual 3D configuration known as the “Barr body.” The authors constructed a diploid Hi-C map of human GM12878 cells and showed that the Xi chromosome had a distinctive superstructure. It contained superdomains, which are unusually large contact domains. It also contained unusually large chromatin loops called “superloops.” Both superdomains and superloops can span dozens of megabases of the genome. Superloop anchors, like the macrosatellite repeat DXZ4, tended to colocate simultaneously. Deleting DXZ4 on Xi led to the disappearance of superdomains and superloops, changes in compartmentalization patterns, and changes in the distribution of chromatin marks.

Illumina Technology: HiSeq 2000 System

Krijger PH, Di Stefano B, de Wit E, et al. Cell-of-Origin-Specific 3D Genome Structure Acquired during Somatic Cell Reprogramming. *Cell Stem Cell*. 2016;18:597-610.

The authors used Hi-C contact maps for each of 4 founder cell types and their respective p3 and p20 iPSC derivatives to investigate how nuclear organization changes during reprogramming. They found that early passage iPSCs carried topological hallmarks that enabled recognition of their cell of origin. These hallmarks were not remnants of somatic chromosome topologies. Instead, the distinguishing topological features were acquired during reprogramming.

Illumina Technology: HiSeq 2000 System

Veluchamy A, Jegu T, Ariel F, et al. LHP1 Regulates H3K27me3 Spreading and Shapes the Three-Dimensional Conformation of the Arabidopsis Genome. *PLoS One*. 2016;11:e0158936.

Like Heterochromatin Protein 1 (LHP1) controls actively transcribed genes and is a member of the plant-specific polycomb-group (PcG) family originally identified in Drosophila. The authors used Hi-C to map the spatial contacts and distribution of genes in chromatin between different parts of the Arabidopsis genome for wild-type and lhp1 plants. Chromosomal contact maps at 100 kb resolution showed significant changes between wt and lhp1 plants. Additional experiments showed that LHP1 was responsible for the spreading of H3K27me3 toward the 3’ end of the gene body. The authors also identified a subset of LHP1-activated genes that shape local chromatin topology to control transcriptional coregulation.

Illumina technology: HiSeq 2500 System

Associated Kits

TruSeq Nano DNA Library Prep Kit

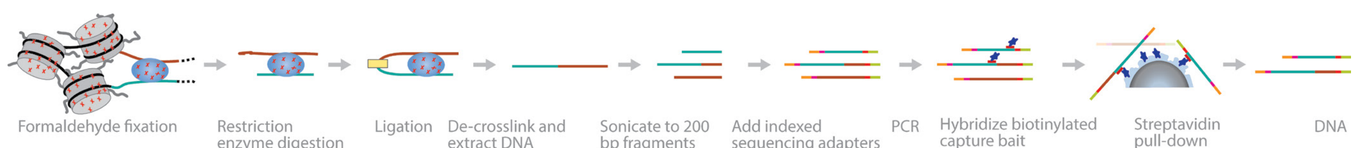
TruSeq DNA Sample Prep Kit

TruSeq DNA PCR-Free Library Prep Kit

NG Capture-C: Next-Generation Capture-C

NG Capture-C is a refinement of 3C-Seq[146] and Hi-C.[147] It represents a family of methods used to analyze chromatin interactions. NG Capture-C adds multiple pull-down steps of the biotinylated fragments with magnetic beads to the 3C method.

The protocol uses formaldehyde fixation followed by restriction enzyme digestion and ligation to form ~10 kb concatamers. The DNA is extracted and sonicated. Indexing adapters are added, and the samples are pooled, purified by pull-down, and PCR-amplified. The pull-down and PCR steps can be repeated to yield up to a million-fold enrichment.[148]



A schematic overview of NG Capture-C.

Table 7. Advantages and Disadvantages of NG Capture-C

Advantages	Disadvantages
• High sensitivity to detect cis and trans interactions	• None reported
• Low sample input requirements	• None reported
• Sonicated capture fragments reduces cost compared to capture-C	• None reported
• Sonicated capture fragments act as UMIs to reduce PCR bias	• None reported

Reviews

Sati S and Cavalli G Chromosome conformation capture technologies and their impact in understanding genome function. *Chromosoma*. 2016;

References

Davies JO, Telenius JM, McGowan SJ, et al. Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nat Methods*. 2016;13:74-80.

The authors used NG Capture-C to detect ligation junctions, equivalent to the detection of interactions present in 1 in 10,000 cells, at single-restriction-fragment resolution (~250 bp). This result exceeds the sensitivity and resolution of detection of current 3C methods and complementary methods, such as FISH.

Illumina technology: HiSeq System, MiSeq System

Hay D, Hughes JR, Babbs C, et al. Genetic dissection of the alpha-globin super-enhancer in vivo. Nat Genet. 2016;48:895-903.

The authors used homologous recombination to generate 7 mouse models in which each constituent of the proposed α-globin superenhancer was deleted, individually and in informative pairs, to dissect its function. NG Capture-C detected the only statistically significant changes, which occurred at the α-globin promoters.

Illumina technology: NextSeq System

Associated Kits

TruSeq ChIP Library Prep Kit

TruSeq Nano DNA Library Prep Kit

TruSeq DNA Sample Preparation Kit

TruSeq DNA PCR-Free Library Prep Kit

Nextera DNA Library Prep Kit

Nextera XT DNA Library Prep Ki

146. Duan Z, Andronescu M, Schutz K, et al. A genome-wide 3C-method for characterizing the three-dimensional architectures of genomes. *Methods*. 2012;58:277-288.

147. Lieberman-Aiden E, van Berkum NL, Williams L, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 2009;326:289-293.

148. Davies JO, Telenius JM, McGowan SJ, et al. Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nat Methods*. 2016;13:74-80.

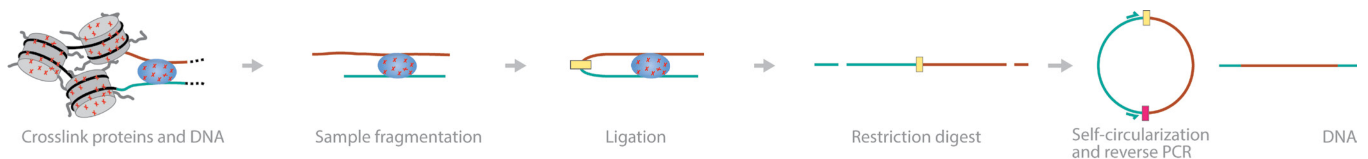
149. Zhao Z, Tavoosidana G, Sjolinder M, et al. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet*. 2006;38:1341-1347.

150. Sajan SA and Hawkins RD Methods for identifying higher-order chromatin structure. *Annu Rev Genomics Hum Genet*. 2012; 13:59-82

4C-seq: Circular Chromatin Conformation Capture

4C[149], also called 4C-seq, is a method similar to 3C and is sometimes called circular 3C. It allows the unbiased detection of all genomic regions that interact with a particular region of interest.[150]

In this method, DNA-protein complexes are crosslinked using formaldehyde. The sample is fragmented, and the DNA is ligated and digested. The resulting DNA fragments self-circularize, followed by reverse PCR and sequencing. Deep sequencing provides base-pair resolution of the ligated fragments.



A schematic overview of 4C.

Table 8. Advantages and Disadvantages of 4C

Advantages	Disadvantages
• Preferred strategy to assess the DNA contact profile of individual genomic sites • Highly reproducible data	• Will miss local interactions (< 50 kb) from the region of interest • Large circles do not amplify efficiently

Reviews

Acemel RD, Tena JJ, Irastorza-Azcarate I, Marletaz F, Gomez-Marin C, et al. A single three-dimensional chromatin compartment in amphioxus indicates a stepwise evolution of vertebrate Hox bimodal regulation. *Nat Genet*. 2016;48:336-341.

De S, Mitra A, Cheng Y, Pfeifer K and Kassis JA Formation of a Polycomb-Domain in the Absence of Strong Polycomb Response Elements. *PLoS Genet*. 2016;12:e1006200.

Eckart N, Song Q, Yang R, et al. Functional Characterization of Schizophrenia-Associated Variation in CACNA1C. *PLoS One*. 2016;11:e0157086.

Kaaij LJ, Mokry M, Zhou M, et al. Enhancers reside in a unique epigenetic environment during early zebrafish development. *Genome Biol*. 2016;17:146.

Kandaswamy R, Sava GP, Speedy HE, et al. Genetic Predisposition to Chronic Lymphocytic Leukemia Is Mediated by a BMF Super-Enhancer Polymorphism. *Cell Rep*. 2016;16:2061-2067.

Lin CY, Erkek S, Tong Y, et al. Active medulloblastoma enhancers reveal subgroup-specific cellular origins. *Nature*. 2016;530:57-62.

Proudhon C, Snetkova V, Raviram R, et al. Active and Inactive Enhancers Cooperate to Exert Localized and Long-Range Control of Gene Regulation. *Cell Rep*. 2016;15:2159-2169.

Rocha PP, Raviram R, Fu Y, Kim J, Luo VM, et al. A Damage-Independent Role for 53BP1 that Impacts Break Order and Igh Architecture during Class Switch Recombination. *Cell Rep*. 2016;16:48-55.

Vermunt MW, Tan SC, Castelijns B, et al. Epigenomic annotation of gene regulatory alterations during evolution of the primate brain. *Nat Neurosci*. 2016;19:494-503.

Wang Q, Sawyer IA, Sung MH, et al. Cajal bodies are linked to genome conformation. *Nat Commun*. 2016;7:10966.

Wani AH, Boettiger AN, Schorderet P, et al. Chromatin topology is coupled to Polycomb group protein subnuclear organization. *Nat Commun*. 2016;7:10291.

References

Cai M, Kim S, Wang K, Farnham PJ, Coetzee GA and Lu W 4C-seq revealed long-range interactions of a functional enhancer at the 8q24 prostate cancer risk locus. *Scientific reports*. 2016;6:22462.

Genome-wide association studies have identified > 100 independent susceptibility loci for prostate cancer, including a “hot spot” at 8q24. To identify genome-wide partners interacting with this hot spot, the authors coupled 4C-seq to an enhancer at 8q24 as “bait” in cell lines LNCaP and C4-2B. The 4C-identified regions were distributed in open nuclear compartments, featuring active histone marks H3K4me1, H3K4me2, and H3K27Ac.

Illumina technology: HiSeq 2000 System

Loviglio MN, Leleu M, Mannik K, et al. Chromosomal contacts connect loci associated with autism, BMI and head circumference phenotypes.
***Mol Psychiatry*. 2016.**

The authors used 4C-seq to identify chromosomal regions that physically associate with the promoters of genes linked to autism-spectrum and other disorders, including MVP, KCTD13, ALDOA, TBX6 and MAPK3. They found that 2 CNV-prone regions at 16p11.2 were reciprocally engaged in complex chromatin looping, as successfully confirmed by 4C-seq, FISH, and Hi-C. Their results suggest that disruption of chromatin interplays at 16p11.2 could play a role in the observed phenotypes.

Illumina Technology: HiSeq System

Yang R, Kerschner JL, Gosalia N, et al. Differential contribution of cis-regulatory elements to higher order chromatin structure and expression of the CFTR locus. *Nucleic Acids Res*. 2016;44:3082-3094.

The authors used CRISPR/Cas9 editing of cis-regulatory elements and small interfering RNA (siRNA)-mediated depletion of architectural proteins to determine the relative contribution of structural elements and enhancers to the higher order structure and expression of the cystic fibrosis transmembrane conductance regulator (CFTR) locus. They used CRISPR/Cas9-mediated deletion of a CTCF-binding insulator element 5’ to the CFTR locus and a pivotal intronic enhancer, followed by 4C-seq. The results were consistent with a mechanism coordinating regulatory elements across the locus, which senses structural perturbations to maintain normal gene expression. However, they found that the loss of a key intronic enhancer on CFTR transcription could not be rescued by structural changes in the locus.

Illumina technology: HiSeq System

Associated kits

TruSeq ChIP Library Prep Kit

TruSeq Nano DNA Library Prep Kit

TruSeq DNA Sample Prep Kit

TruSeq DNA PCR-Free Library Prep Kit

For Research Use Only. Not for use in diagnostic procedures.

UMI-4C: Circular Chromosome Conformation Capture with Unique Molecular Identifiers

This variation on the 4C approach uses UMIs to derive high-complexity quantitative chromosome contact profiles with controlled signal-to-noise ratios.[151] It is an efficient and accurate method for analyzing targeted loci. The method is paired with software to analyze the data (https://bitbucket.org/tanaylab/umi4cpackage).

A schematic overview of UMI-4C.

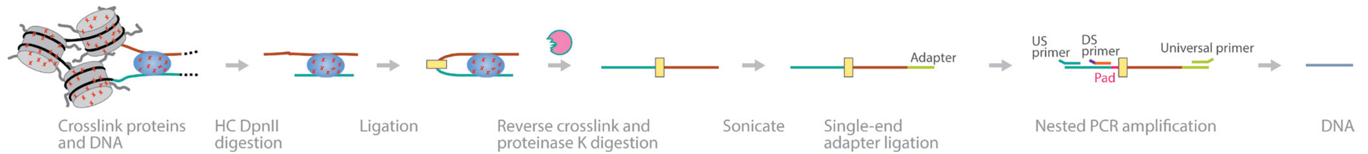


Table 9. Advantages and Disadvantages of UMI-4C

Advantages	Disadvantages
• Improved sensitivity and specificity • Multiplexing allows robust comparison of contact distributions between loci and conditions	• None reported

References

Schwartzman O, Mukamel Z, Oded-Elkayam N, et al. UMI-4C for quantitative and targeted chromosomal contact profiling. *Nat Methods*. 2016;13:685-691.

The authors describe the 4C-UMI method and show that it requires modest sequencing depth (100,000 reads per bait). The method can be multiplexed easily, allowing the selection of multiple viewpoints.

Illumina technology: MiSeq System, NextSeq System, HiSeq System

Associated Kits

TruSeq ChIP Library Prep Kit

TruSeq Nano DNA Library Prep Kit

TruSeq DNA Sample Preparation Kit

TruSeq DNA PCR-Free Library Prep Kit

Nextera DNA Library Prep Kit

Nextera XT DNA Library Prep Kit

5C: Chromatin Conformation Capture Carbon Copy

5C[152] allows concurrent determination of interactions among multiple sequences and is a high-throughput version of 3C.[153]

In this method, DNA-protein complexes are crosslinked using formaldehyde. The sample is fragmented and the DNA ligated and digested with restriction enzymes. The resulting DNA fragments are amplified using ligation-mediated PCR and sequenced. Deep sequencing provides base-pair resolution of the ligated fragments.

A schematic overview of 5C.

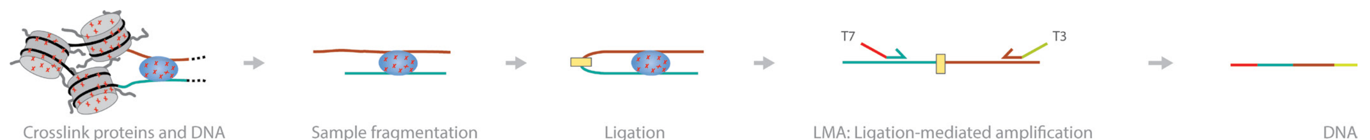


Table 10. Advantages and Disadvantages of UMI-4C

Advantages	Disadvantages
• Different from 4C, 5C provides a matrix of interaction frequencies for many pairs of sites[154] • Can be used to reconstruct the (average) 3D conformation of larger genomic regions[120]	• Requires a *priori* information of the regulatory sites[155] • Detection may not necessarily mean an interaction, resulting from random chromosomal collisions. • Cannot scale to genome-wide studies that would require a large amount of primers

Reviews

Sati S and Cavalli G Chromosome conformation capture technologies and their impact in understanding genome function. *Chromosoma*. 2016;

Reuter JA, Spacek DV and Snyder MP High-throughput sequencing technologies. *Mol Cell*. 2015; 58:586-597

References

Smith EM, Lajoie BR, Jain G and Dekker J Invariant TAD Boundaries Constrain Cell-Type-Specific Looping Interactions between Promoters and Distal Elements around the CFTR Locus. *Am J Hum Genet*. 2016;98:185-201.

Globally, chromosomes are organized into active and inactive compartments while, at the gene level, looping interactions connect promoters to regulatory elements. Topologically associating domains (TADs) represent an intermediate level of organization. The authors designed a 5C experiment to interrogate looping interactions between HindIII fragments containing TSSs and any other HindIII restriction fragments (distal fragments) in the target region. Their results showed that the same TAD boundaries were present in all cell types, and they suggest that TADs represent a universal chromosome architecture.

Illumina technology: Genome Analyzer IIx System

Associated Kits

TruSeq Nano DNA Library Prep Kit

TruSeq DNA Sample Prep Kit

TruSeq DNA PCR-Free Library Prep Kit

151. Schwartzman O, Mukamel Z, Oded-Elkayam N, et al. UMI-4C for quantitative and targeted chromosomal contact profiling. *Nat Methods*. 2016;13:685-691
152. Dostie J and Dekker J Mapping networks of physical interactions between genomic elements using 5C technology. *Nat Protoc*. 2007;2:988-1002.
153. Sajan SA and Hawkins RD Methods for identifying higher-order chromatin structure. *Annu Rev Genomics Hum Genet*. 2012;13:59-82.
154. De Wit E and de Laat W A decade of 3C technologies: insights into nuclear organization. *Genes Dev*. 2012;26:11-24.
155. Sati S and Cavalli G Chromosome conformation capture technologies and their impact in understanding genome function. *Chromosoma*. 2016;

Genotyping by sequencing

RAD and PE RAD-Seq: Restriction-Site Associated DNA Sequencing

RAD-Seq is a protocol for genotyping and discovery of single-nucleotide polymorphisms (SNPs).[156] This approach is particularly useful for genotyping when a reference genome is not available, such as in ecological studies.[157] PE RAD-Seq, also called RAD-PE, is the same protocol as RAD but uses paired-end sequencing for improved alignments.[158] Several variations, such as ddRADseq,[159] 2b-RAD,[160] SLAF-seq,[161] and hyRAD[162] have been developed to address specific applications, and multiple software packages are available to analyze RAD data.[163,164]

A schematic overview of RAD-Seq.

Table 11. Advantages and Disadvantages of RAD-Seq.

Advantages	Disadvantages
• No reference genome required[165]	• There can be gaps in the genome coverage
• Relatively inexpensive, compared to whole-genome sequencing	• Requires high-quality DNA (see hyRAD for low-quality DNA)
• The degree of genome coverage can be adjusted by selecting various restriction enzymes and fragment sizes	• Sequence polymorphism at the DNA restriction sites causes a progressive loss of shared restriction sites among diverging clades

Reviews

Andrews KR, Good JM, Miller MR, Luikart G and Hohenlohe PA Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet. 2016;17:81-92.

da Fonseca RR, Albrechtsen A, Themudo GE, et al. Next-generation biology: Sequencing and data analysis approaches for non-model organisms. *Mar Genomics*. 2016;30:3-13.

Kagale S, Koh C, Clarke WE, Bollina V, Parkin IA and Sharpe AG Analysis of Genotyping-by-Sequencing (GBS) Data. *Methods Mol Biol*. 2016;1374:269-284.

Kim C, Guo H, Kong W, Chandnani R, Shuang LS and Paterson AH Application of genotyping by sequencing technology to a variety of crop breeding programs. *Plant science : an international journal of experimental plant biology*. 2016;242:14-22.

Manel S, Perrier C, Pratlong M, et al. Genomic resources and their influence on the detection of the signal of positive selection in genome scans. *Mol Ecol*. 2016;25:170-184.

Sanders IR and Rodriguez A Aligning molecular studies of mycorrhizal fungal diversity with ecologically important levels of diversity in ecosystems. *ISME J*. 2016;10:2780-2786.

Bian C, Hu Y, Ravi V, et al. The Asian arowana (Scleropages formosus) genome provides new insights into the evolution of an early lineage of teleosts. *Scientific reports*. 2016;6:24501.

Clark LV and Sacks EJ TagDigger: user-friendly extraction of read counts from GBS and RAD-seq data. Source code for biology and medicine. 2016;11:11, Hou Y, Nowak MD, Mirre V, Bjora CS, Brochmann C and Popp M RAD-seq data point to a northern origin of the arctic-alpine genus Cassiope (Ericaceae). *Molecular phylogenetics and evolution*. 2015;95:152-160.

Diaz-Arce N, Arrizabalaga H, Murua H, Irigoien X and Rodriguez-Ezpeleta N RAD-seq derived genome-wide nuclear markers resolve the phylogeny of tunas. *Mol Phylogenet Evol*. 2016;102:202-207.

Fan W, Zong J, Luo Z, et al. Development of a RAD-Seq Based DNA Polymorphism Identification Software, AgroMarker Finder, and Its Application in Rice Marker-Assisted Breeding. *PLoS One*. 2016;11:e0147187.

Fernandez R, Schubert M, Vargas-Velazquez AM, et al. A genomewide catalogue of single nucleotide polymorphisms in white-beaked and Atlantic white-sided dolphins. *Mol Ecol Resour*. 2016;16:266-276

Herrera S and Shank TM RAD sequencing enables unprecedented phylogenetic resolution and objective species delimitation in recalcitrant divergent taxa. *Mol Phylogenet Evol*. 2016;100:70-79

Jiang N, Zhang F, Wu J, et al. A highly robust and optimized sequence-based approach for genetic polymorphism discovery and genotyping in large plant populations. *Theor Appl Genet*. 2016;129: 1739-1757.

Manthey JD and Robbins MB Genomic insights into hybridization in a localized region of sympatry between pewee sister species (Contopus sordidulus × C. virens) and their chromosomal patterns of differentiation. *Avian Research*. 2016;7:

Qiu S, Bergero R, Guirao-Rico S, et al. RAD mapping reveals an evolving, polymorphic and fuzzy boundary of a plant pseudoautosomal region. *Mol Ecol*. 2016;25:414-430.

Ravinet M, Westram A, Johannesson K, Butlin R, Andre C and Panova M Shared and nonshared genomic divergence in parallel ecotypes of Littorina saxatilis at a local scale. *Mol Ecol.* 2016;25: 287-305.

Takahashi T and Sota T A robust phylogeny among major lineages of the East African cichlids. Mol *Phylogenet Evol*. 2016;100:234-242.

Yang H, Wei CL, Liu HW, et al. Genetic Divergence between Camellia sinensis and Its Wild Relatives Revealed via Genome-Wide SNPs from RAD Sequencing. *PLoS One*. 2016;11:e0151424

Yang J, Guo B, Shikano T, Liu X and Merila J Quantitative trait locus analysis of body shape divergence in nine-spined sticklebacks based on high-density SNP-panel. *Scientific reports*. 2016;6:26632

Zhao Y, Zhang C, Chen H, et al. QTL mapping for bacterial wilt resistance in peanut (Arachis hypogaea L.). *Mol Breed*. 2016;36:13

Combosch DJ and Vollmer SV Trans-Pacific RAD-Seq population genomics confirms introgressive hybridization in Eastern Pacific Pocillopora corals. *Molecular phylogenetics and evolution*. 2015;

DaCosta JM and Sorenson MD ddRAD-seq phylogenetics based on nucleotide, indel, and presence-absence polymorphisms: Analyses of two avian genera with contrasting histories. *Molecular phylogenetics and evolution.* 2015;

Demos TC, Kerbis Peterhans JC, Joseph TA, Robinson JD, Agwanda B and Hickerson MJ Comparative Population Genomics of African Montane Forest Mammals Support Population Persistence across a Climatic Gradient and Quaternary Climatic Cycles. *PloS one*. 2015;10:e0131800

Guo B, DeFaveri J, Sotelo G, Nair A and Merila J Population genomic evidence for adaptive differentiation in Baltic Sea three-spined sticklebacks. *BMC Biol.* 2015;13:19.

Liu S, Clark LV, Swaminathan K, Gifford JM, Juvik JA and Sacks EJ High-density genetic map ofMiscanthus sinensisreveals inheritance of zebra stripe. *GCB Bioenergy*. 2016;8:616-630

Longo G and Bernardi G The evolutionary history of the embiotocid surfperch radiation based on genome-wide RAD sequence data. *Molecular phylogenetics and evolution*. 2015;88:55-63

Marubodee R, Ogiso-Tanaka E, Isemura T, et al. Construction of an SSR and RAD-Marker Based Molecular Linkage Map of Vigna vexillata (L.) A. Rich. *PloS one*. 2015;10:e0138942

Mathers TC, Hammond RL, Jenner RA, Hanfling B, Atkins J and Gomez A Transition in sexual system and sex chromosome evolution in the tadpole shrimp Triops cancriformis. Heredity. 2015;

Shao C, Niu Y, Rastas P, et al. Genome-wide SNP identification for the construction of a high-resolution genetic map of Japanese flounder (Paralichthys olivaceus): applications to QTL mapping of Vibrio anguillarum disease resistance and comparative genomic analysis. *DNA research : an international journal for rapid publication of reports on genes and genomes*. 2015;

Takahashi T and Moreno E A RAD-based phylogenetics for Orestias fishes from Lake Titicaca. *Molecular phylogenetics and evolution*. 2015;

Tennessen JA, Bonner KM, Bollmann SR, et al. Genome-Wide Scan and Test of Candidate Genes in the Snail Biomphalaria glabrata Reveal New Locus Influencing Resistance to Schistosoma mansoni. *PLoS neglected tropical diseases*. 2015;9:e0004077.

Yu S, Chu W, Zhang L, et al. Identification of Laying-Related SNP Markers in Geese Using RAD Sequencing. *PloS one*. 2015;10:e0131572.

Zhou Z, Liu S, Dong Y, et al. High-Density Genetic Mapping with Interspecific Hybrids of Two Sea Urchins, Strongylocentrotus nudus and S. intermedius, by RAD Sequencing. *PloS one*. 2015;10:e0138585.

156. Baird NA, Etter PD, Atwood TS, et al. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*. 2008;3:e3376.
157. Andrews KR, Good JM, Miller MR, Luikart G and Hohenlohe PA Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat Rev Genet*. 2016;17:81-92.
158. Willing EM, Hoffmann M, Klein JD, Weigel D and Dreyer C Paired-end RAD-seq for de novo assembly and marker design without available reference. *Bioinformatics*. 2011;27:2187-2193.
159. Peterson BK, Weber JN, Kay EH, Fisher HS and Hoekstra HE Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One*. 2012;7:e37135.
160. Wang S, Meyer E, McKay JK and Matz MV 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nat Methods*. 2012;9:808-810.
161. Sun X, Liu D, Zhang X, et al. SLAF-seq: an efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. *PLoS One*. 2013;8:e58700
162. Suchan T, Pitteloud C, Gerasimova NS, et al. Hybridization Capture Using RAD Probes (hyRAD), a New Tool for Performing Genomic Analyses on Collection Specimens. *PLoS One*. 2016;11:e0151651
163. Fan W, Zong J, Luo Z, et al. Development of a RAD-Seq Based DNA Polymorphism Identification Software, AgroMarker Finder, and Its Application in Rice Marker-Assisted Breeding. *PLoS One*. 2016;11:e0147187.
164. Catchen J, Hohenlohe PA, Bassham S, Amores A and Cresko WA Stacks: an analysis tool set for population genomics. *Mol Ecol*. 2013;22:3124-3140.
165. Reitzel AM, Herrera S, Layden MJ, Martindale MQ and Shank TM Going where traditional markers have not gone before: utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. *Mol Ecol*. 2013;22:2953-2970.
166. Suchan T, Pitteloud C, Gerasimova NS, et al. Hybridization Capture Using RAD Probes (hyRAD), a New Tool for Performing Genomic Analyses on Collection Specimens. *PLoS One*. 2016;11:e0151651

For Research Use Only. Not for use in diagnostic procedures.

References

Ren P, Peng W, You W, et al. Genetic mapping and quantitative trait loci analysis of growth-related traits in the small abalone Haliotis diversicolor using restriction-site-associated DNA sequencing. *Aquaculture*. 2016;454:163-170.

The authors used RAD-Seq to construct a high-resolution linkage map of the abalone *Haliotis diversicolor* for growth-related quantitative trait locus (QTL) analysis. They were able to build a reduced-representation library with 3756 loci in more than 95% of the offspring. Based on this map, they identified 15
QTLs for 6 growth-related traits.

Illumina technology: HiSeq 2000 System

Wang J, Xue DX, Zhang BD, Li YL, Liu BJ and Liu JX Genome-Wide SNP Discovery, Genotyping and Their Preliminary Applications for Population Genetic Inference in Spotted Sea Bass (Lateolabrax maculatus). *PLoS One*. 2016;11:e0157809.

The researchers used PE RAD-Seq on 30 individuals from 2 populations to discover 22,648 SNPs across the genome of *L. maculatus*. The results showed shallow, but significant, genetic differentiation between the 2 populations.

Illumina technology: HiSeq 2000 System

He T, D'Agui H, Lim SL, Enright NJ and Luo Y Evolutionary potential and adaptation of Banksia attenuata (Proteaceae) to climate and fire regime in southwestern Australia, a global biodiversity hotspot. *Scientific reports*. 2016;6:26315.

This study applied RAD-Seq and environmental association analysis to 80 plants and found candidate genes associated with rainfall gradients, temperatures, and fire intervals. The authors discovered that overall population adaptive genetic variation was affected significantly by shortened fire intervals, whereas declining rainfall and rising temperature did not have a detectable influence. Gene annotation further revealed 4 genes with functions in stress tolerance, the regulation of stomatal opening and closure, energy use, and morphogenesis with adaptation to climate and fire intervals.

Illumina technology: HiSeq 2000 System

Associated Kits

TruSeq Nano DNA Library Prep Kit

TruSeq DNA PCR-Free Library Prep Kit

167. Peterson BK, Weber JN, Kay EH, Fisher HS and Hoekstra HE Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One*. 2012;7:e37135.
168. Baird NA, Etter PD, Atwood TS, et al. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*. 2008;3:e3376.
169. Willing EM, Hoffmann M, Klein JD, Weigel D and Dreyer C Paired-end RAD-seq for de novo assembly and marker design without available reference. *Bioinformatics*. 2011;27:2187-2193.
170. Wang S, Meyer E, McKay JK and Matz MV 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nat Methods*. 2012;9:808-810.
171. Sun X, Liu D, Zhang X, et al. SLAF-seq: an efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. *PLoS One*. 2013;8:e58700.
172. Suchan T, Pitteloud C, Gerasimova NS, et al. Hybridization Capture Using RAD Probes (hyRAD), a New Tool for Performing Genomic Analyses on Collection Specimens. *PLoS One*. 2016;11:e0151651.
173. Fan W, Zong J, Luo Z, et al. Development of a RAD-Seq Based DNA Polymorphism Identification Software, AgroMarker Finder, and Its Application in Rice Marker-Assisted Breeding. *PLoS One*. 2016;11:e0147187.
174. Catchen J, Hohenlohe PA, Bassham S, Amores A and Cresko WA Stacks: an analysis tool set for population genomics. *Mol Ecol*. 2013;22:3124-3140
175. Reitzel AM, Herrera S, Layden MJ, Martindale MQ and Shank TM Going where traditional markers have not gone before: utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. *Mol Ecol*. 2013;22:2953-2970.
176. Suchan T, Pitteloud C, Gerasimova NS, et al. Hybridization Capture Using RAD Probes (hyRAD), a New Tool for Performing Genomic Analyses on Collection Specimens. *PLoS One*. 2016;11:e0151651

ddRADseq: Double Digest Restriction-Site Associated DNA Marker Generation

ddRADseq,[167] also called ddRAD, is a variation on the RAD sequencing protocol,[168] which is used for SNP discovery and genotyping.[169] In this variation, the fragment shearing is replaced with a second restriction digestion to improve the tunability and accuracy of the size-selection step. The protocol also includes a second index to allow combinatorial indexing. Several RAD variations, such as 2b-RAD,[170] SLAF-seq,[171] and hyRAD,[172] have been developed to address specific applications, and multiple software packages are available to analyze RAD data.[173,174]

In this method, gDNA is first digested with a restriction enzyme, and a barcoded P1 adapter is ligated to the fragments. The adapter-ligated fragments from different samples are combined, if samples are multiplexed, and the DNA is digested by a second restriction enzyme. The fragments are size-selected and purified. The P2 adapter-primers are ligated, and the fragments are amplified to produce the sequencing library.

A schematic overview of ddRADseq.

Table 12. Advantages and Disadvantages of ddRADseq

Advantages	Disadvantages
• No reference genome required[175] • Relatively inexpensive, compared to whole-genome sequencing. The degree of genome coverage can be adjusted by selecting various restriction enzymes	• There can be gaps in the genome coverage • Requires high-quality DNA (see hyRAD for ow-quality DNA)

Reviews

Andrews KR, Good JM, Miller MR, Luikart G and Hohenlohe PA Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat Rev Genet*. 2016;17:81-92.

Kagale S, Koh C, Clarke WE, Bollina V, Parkin IA, et al. Analysis of Genotyping-by-Sequencing (GBS) Data. *Methods Mol Biol*. 2016;1374:269-284.

Brown JK, Taggart JB, Bekaert M, et al. Mapping the sex determination locus in the hapuku (Polyprion oxygeneios) using ddRAD sequencing. *BMC Genomics*. 2016;17:448.

Clark LV and Sacks EJ TagDigger: user-friendly extraction of read counts from GBS and RAD-seq data. *Source code for biology and medicine*. 2016;11:11.

Hou Y, Nowak MD, Mirre V, Bjora CS, Brochmann C and Popp M RAD-seq data point to a northern origin of the arctic-alpine genus Cassiope (Ericaceae). *Mol Phylogenet Evol*. 2016;95:152-160.

Shirasawa K, Hirakawa H and Isobe S Analytical workflow of double-digest restriction site-associated DNA sequencing based on empirical and in silico optimization in tomato. *DNA Res*. 2016;23:145-153.

Wu Z, Wang B, Chen X, et al. Evaluation of Linkage Disequilibrium Pattern and Association Study on Seed Oil Content in Brassica napus Using ddRAD Sequencing. *PLoS One*. 2016;11:e0146383.

Yang GQ, Chen YM, Wang JP, et al. Development of a universal and simplified ddRAD library preparation approach for SNP discovery and genotyping in angiosperm plants. *Plant Methods*. 2016;12:39

Cai G, Yang Q, Yi B, et al. A bi-filtering method for processing single nucleotide polymorphism array data improves the quality of genetic map and accuracy of quantitative trait locus mapping in doubled haploid populations of polyploid Brassica napus. *BMC Genomics*. 2015;16:409

Davik J, Sargent DJ, Brurberg MB, Lien S, Kent M and Alsheikh M A ddRAD Based Linkage Map of the Cultivated Strawberry, Fragaria xananassa. *PLoS One*. 2015;10:e0137746

Leache AD, Banbury BL, Felsenstein J, de Oca AN and Stamatakis A Short Tree, Long Tree, Right Tree, Wrong Tree: New Acquisition Bias Corrections for Inferring SNP Phylogenies. *Systematic biology*. 2015;64:1032-1047

Leache AD, Chavez AS, Jones LN, Grummer JA, Gottscho AD and Linkem CW Phylogenomics of phrynosomatid lizards: conflicting signals from sequence capture versus restriction site associated DNA sequencing. *Genome Biol Evol*. 2015;7:706-719

Meik JM, Streicher JW, Lawing AM, Flores-Villela O and Fujita MK Limitations of climatic data for inferring species boundaries: insights from speckled rattlesnakes. *PLoS One*. 2015;10:e0131435

References

DaCosta JM and Sorenson MD ddRAD-seq phylogenetics based on nucleotide, indel, and presence-absence polymorphisms: Analyses of two avian genera with contrasting histories. *Mol Phylogenet Evol*. 2016;94:122-135.

This study used ddRADseq for phylogenetic analyses of the *Lagonosticta* firefinches (family Estrildidae) and the *Vidua* brood parasitic finches (family Viduidae). The researchers obtained > 1000 homologous loci despite ~20 million years divergence. In addition to nucleotide polymorphisms, the ddRADseq data yielded large sets of indel and locus presence–absence polymorphisms, all of which had higher consistency indices than mitochondrial DNA (mtDNA) sequence data.

Illumina technology: HiSeq 2000 System

Lal MM, Southgate PC, Jerry DR and Zenger KR Fishing for divergence in a sea of connectivity: The utility of ddRADseq genotyping in a marine invertebrate, the black-lip pearl oyster Pinctada margaritifera. *Mar Genomics*. 2016;25:57-68.

The authors applied ddRADseq to *Pinctada margaritifera* and detected 5243 high-quality, genome-wide SNP markers. They were able to assess population structure, genome diversity, and perform association testing in 156 individuals belonging to 3 wild populations and 1 hatchery-produced population from the Fiji Islands. They also found shallow, but significant, population structure among the wild populations.

Illumina technology: HiSeq 2000 System

Associated Kits

TruSeq Nano DNA Library Prep Kit

TruSeq DNA PCR-Free Library Prep Kit

2b-RAD: RAD with Type IIB Restriction Endonucleases

2b-RAD is similar to ddRadseq but uses type IIB restriction enzymes (BsaXI or AlfI), which will cleave upstream and downstream of a recognition site. This shears the target genome into a large number of DNA fragments with a constant length of 33 bp (BsaXI) or 36 bp (AlfI). These short DNA fragments can be sequenced to determine genetic variants.

In this method, gDNA is first digested with a restriction enzyme (BsaXI), and adapters with partial (NNN) overhangs are ligated to the fragments. The adapter-ligated fragments from different samples are combined, and the fragments are amplified to produce the sequencing library.

A schematic overview of 2b-RAD.

Table 13. Advantages and Disadvantages of 2b-RAD.

Advantages	Disadvantages
• Highly reduced 2b-RAD libraries require much less sequencing for accurate genotyping • High density of markers. • No interim purification steps, reducing losses and processing time	• Requires a reference genome • Short tags may not be long enough for efficient locus discrimination in complex genomes

Reviews

Andrews KR, Good JM, Miller MR, Luikart G and Hohenlohe PA Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat Rev Genet*. 2016;

da Fonseca RR, Albrechtsen A, Themudo GE, Ramos-Madrigal J, Sibbesen JA, et al. Next-generation biology: Sequencing and data analysis approaches for non-model organisms. *Mar Genomics*. 2016;30:3-13.

Jiang N, Zhang F, Wu J, Chen Y, Hu X, et al. A highly robust and optimized sequence-based approach for genetic polymorphism discovery and genotyping in large plant populations. *Theor Appl Genet*. 2016;129:1739-1757.

Kagale S, Koh C, Clarke WE, Bollina V, Parkin IA, et al. Analysis of Genotyping-by-Sequencing (GBS) Data. *Methods Mol Biol.* 2016;1374:269-284.

Manel S, Perrier C, Pratlong M, Abi-Rached L, Paganini J, et al. Genomic resources and their influence on the detection of the signal of positive selection in genome scans. *Mol Ecol*. 2016;25:170-184.

References

Pecoraro C, Babbucci M, Villamor A, et al. Methodological assessment of 2b-RAD genotyping technique for population structure inferences in yellowfin tuna (Thunnus albacares). *Mar Genomics*. 2016;25:43-48.

The global population genetic structure of yellowfin tuna (*Thunnus albacares*) is poorly understood. The authors used 2b-RAD to show shallow but significant population structure among oceans (fixation index (F_{ST}) = 0.0273; $P < 0.01$). Discriminant analysis of principal components supported the presence of genetically discrete yellowfin tuna populations among 3 oceanic pools.

Illumina Technology: HiSeq 2500 System

Fu B, Liu H, Yu X and Tong J A high-density genetic map and growth related QTL mapping in bighead carp (Hypophthalmichthys nobilis).
Scientific reports. 2016;6:28679.

The authors used 2b-RAD sequencing of 117 individuals in an F1 family to construct a map with 3121 SNP markers. Based on this genetic map, they identified
1 genome-wide significant and 37 suggestive QTLs for 5 growth-related traits in 6 linkage groups (LG3, LG11, LG15, LG18, LG19, LG22).

Illumina technology: HiSeq 2500 System

Associated Kits

TruSeq DNA PCR-Free Library Prep Kit

SLAF-seq: Specific Locus Amplified Fragment Sequencing

SLAF-seq is an optimized version of ddRADseq, specifically intended for large-scale genotyping experiments.[177]

The enzymes and the sizes of the restriction fragments are optimized with training data to ensure even distribution and avoid repeats. The fragments are also selected over a tight range, to optimize PCR amplification. The protocol is similar to ddRAD, with a first digestion with MseI, heat inactivation, and a second digestion with AluI. The resulting fragments are PCR-amplified, adapters are added, and the fragments are purified to produce the sequencing library.

A schematic overview of SLAF-seq.

Table 14. Advantages and Disadvantages of SLAF-seq

Advantages	Disadvantages
• Deep sequencing for genotyping accuracy. • Reduced-representation strategy to reduce sequencing costs. • Predesigned reduced-representation scheme to optimize marker efficiency. • Double barcode system for large populations.	• Does not cover the whole genome

Reviews

Andrews KR, Good JM, Miller MR, Luikart G and Hohenlohe PA Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat Rev Gene*t. 2016;

Wei QZ, Fu WY, Wang YZ, et al. Rapid identification of fruit length loci in cucumber (Cucumis sativus L.) using next-generation sequencing (NGS)-based QTL analysis. *Scientific reports*. 2016;6:27496

Xu X, Chao J, Cheng X, et al. Mapping of a Novel Race Specific Resistance Gene to Phytophthora Root Rot of Pepper (Capsicum annuum) Using Bulked Segregant Analysis Combined with Specific Length Amplified Fragment Sequencing Strategy. *PLoS One*. 2016;11:e0151401

Ye Y, Cai M, Ju Y, et al. Identification and Validation of SNP Markers Linked to Dwarf Traits Using SLAF-Seq Technology in Lagerstroemia. *PLoS One*. 2016;11:e0158970

Zhang H, Yi H, Wu M, et al. Mapping the Flavor Contributing Traits on “Fengwei Melon” (Cucumis melo L.) Chromosomes Using Parent Resequencing and Super Bulked-Segregant Analysis. *PLoS One*. 2016;11:e0148150

Cai C, Cheng FY, Wu J, Zhong Y and Liu G The First High-Density Genetic Map Construction in Tree Peony (Paeonia Sect. Moutan) using Genotyping by Specific-Locus Amplified Fragment Sequencing. *PLoS One*. 2015;10:e0128584

Ma JQ, Huang L, Ma CL, et al. Large-Scale SNP Discovery and Genotyping for Constructing a High-Density Genetic Map of Tea Plant Using Specific-Locus Amplified Fragment Sequencing (SLAF-seq). *PLoS One*. 2015;10:e0128798

Qin D, Dong J, Xu F, et al. Characterization and fine mapping of a novel barley Stage Green-Revertible Albino Gene (HvSGRA) by Bulked Segregant Analysis based on SSR assay and Specific Length Amplified Fragment Sequencing. *BMC Genomics*. 2015;16:838

Shan T, Pang S, Li J, Li X and Su L Construction of a high-density genetic map and mapping of a sex-linked locus for the brown alga Undaria pinnatifida (Phaeophyceae) based on large scale marker development by specific length amplified fragment (SLAF) sequencing. *BMC Genomics*. 2015;16:902.

Wang J, Zhang K, Zhang X, et al. Construction of Commercial Sweet Cherry Linkage Maps and QTL Analysis for Trunk Diameter. *PLoS One*. 2015;10:e0141261.

Wang W, Zhang T, Zhang G, et al. Genome-wide association study of antibody level response to NDV and IBV in Jinghai yellow chicken based on SLAF-seq technology. *Journal of applied genetics*. 2015;56:365-373.

Xu F, Sun X, Chen Y, Huang Y, Tong C and Bao J Rapid identification of major QTLs associated with rice grain weight and their utilization. *PLoS One*. 2015;10:e0122206.

Xu X, Lu L, Zhu B, Xu Q, Qi X and Chen X QTL mapping of cucumber fruit flesh thickness by SLAF-seq. *Scientific reports*. 2015;5:15829.

Xu Y, Huang L, Ji D, Chen C, Zheng H and Xie C Construction of a dense genetic linkage map and mapping quantitative trait loci for economic traits of a doubled haploid population of Pyropia haitanensis (Bangiales, Rhodophyta). *BMC Plant Biol*. 2015;15:228.

Zhang J, Zhang Q, Cheng T, et al. High-density genetic map construction and identification of a locus controlling weeping trait in an ornamental woody plant (Prunus mume Sieb. et Zucc). *DNA Res*. 2015; 22:183-191.

Zhang Y, Zhang J, Huang L, et al. A high-density genetic map for P genome of Agropyron Gaertn. based on specific-locus amplified fragment sequencing (SLAF-seq). *Planta*. 2015;242:1335-1347.

References

Zhao X, Huang L, Zhang X, et al. Construction of high-density genetic linkage map and identification of flowering-time QTLs in orchardgrass using SSRs and SLAF-seq. *Scientific reports*. 2016;6:29345.

The authors developed 447,177 markers based on SLAF-seq and used them to perform a comparative genomics analysis between perennial ryegrass and orchardgrass. They found 11 potentially significant QTLs for 2 target traits—heading date (HD) and flowering time (FT).

Illumina technology: HiSeq 2500 System

Geng X, Jiang C, Yang J, Wang L, Wu X and Wei W Rapid Identification of Candidate Genes for Seed Weight Using the SLAF-Seq Method in Brassica napus. *PLoS One*. 2016;11:e0147580.

The researchers used SLAF-seq with association analysis and bulked segregant analysis to identify candidate genes in 1000-seed weight (TSW) data. They found a total of 1933 high-quality polymorphic SLAF markers and 4 TSW-associated markers.

Illumina technology: HiSeq 2500 System

Zhang J, Yuan H, Li M, et al. A High-Density Genetic Map of Tetraploid Salix matsudana Using Specific Length Amplified Fragment Sequencing (SLAF-seq). *PLoS One*. 2016;11:e0157777.

The authors created an intraspecific F1 hybrid population by crossing the salt-sensitive “Yanjiang” variety of an arbor tree species (*Salix matsudana*) as the female parent with the salt-tolerant “9901” variety as the male parent. They genotyped this population, along with its parents. Both the parents and offspring were tetraploid, but the authors were able to construct a genetic map with 6737 SLAF markers. Their data will be used to map quantitative trait loci that modulate salt tolerance and resistance in *Salix*.

Illumina technology: HiSeq 2500 System

Associated Kits

TruSeq DNA PCR-Free Library Prep Kit

177. Sun X, Liu D, Zhang X, et al. SLAF-seq: an efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. *PLoS One*. 2013;8:e58700.
178. Suchan T, Pitteloud C, Gerasimova NS, et al. Hybridization Capture Using RAD Probes (hyRAD), a New Tool for Performing Genomic Analyses on Collection Specimens. *PLoS One*. 2016;11:e0151651.
179. Peterson BK, Weber JN, Kay EH, Fisher HS and Hoekstra HE Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One*. 2012;7:e37135.
180. Ali OA, O'Rourke SM, Amish SJ, et al. RAD Capture (Rapture): Flexible and Efficient Sequence-Based Genotyping. *Genetics*. 2016;202:389-400.

hyRAD: Hybridization RAD for Degraded DNA

hyRAD[178] was developed for use on degraded DNA samples, such as those from museum collections. Museum and preserved samples offer a rich source of valuable specimens, but their degraded DNA is unable to sustain the double-digestion and sample fractionation required by ddRAD.[179] To address this limitation, hyRAD uses biotinylated probes and streptavidin-covered beads to capture and enrich the fragments of interest.

The first step in the process is to generate a ddRAD library from high-quality DNA, usually from an extant specimen. The fragments are size-selected and biotinylated. They can now be used as probes for hybridization capture of shotgun or ddRAD libraries.

A schematic overview of hyRAD.

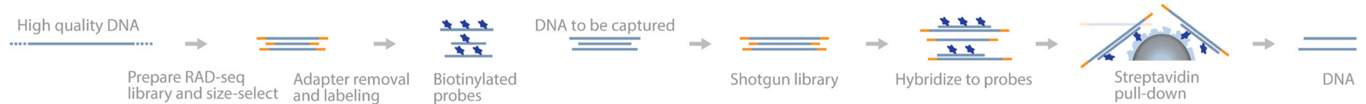


Table 15. Advantages and Disadvantages of hyRAD

Advantages	Disadvantages
• Can be used on degraded DNA. • Can be used on unsequenced genomes.	• Coverage not as complete as with high-quality samples.

Reviews

Andrews KR, Good JM, Miller MR, Luikart G and Hohenlohe PA Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat Rev Genet*. 2016;

Holmes MW, Hammond TT, Wogan GO, et al. Natural history collections as windows on evolutionary processes. *Mol Ecol*. 2016;25:864-881.

References

Suchan T, Pitteloud C, Gerasimova NS, et al. Hybridization Capture Using RAD Probes (hyRAD), a New Tool for Performing Genomic Analyses on Collection Specimens. *PLoS One*. 2016;11:e0151651.

This study describes the hyRAD protocol and demonstrates its performance. The researchers obtained a large set of orthologous loci from fresh and museum samples from a non–model butterfly species. They discovered a high proportion of SNPs present in all 8 analyzed specimens, including 58-year-old museum samples.

Illumina technology: MiSeq System, HiSeq System

Associated Kits

TruSeq DNA PCR-Free Library Prep Kit

Rapture: Restriction-Site Associated DNA Capture

Rapture is a massively parallel, targeted DNA sequencing technique that combines RAD-Seq and sequence capture to compare multiple genes of interest among large numbers of samples.[180] The targeted sequencing is based on identifying restriction enzyme sites specifically near the loci of interest.

gDNA samples are pooled into individual wells in plates and digested with selected restriction enzymes. Biotinylated RAD adapters with well-specific barcodes are ligated to the sticky ends before pooling all the wells from each plate. The barcoded DNA fragments are randomly sheared and bound to streptavidin beads. Again, using restriction enzymes, fragments are cleaved from the streptavidin beads and used in standard DNA library preparation kits with plate-specific barcode labels. Next, libraries from both plates are pooled and hybridized with biotinylated bait-oligos specific to each RAD tag, before a final streptavidin pull-down. The isolated DNA fragments are sequenced and arranged according to their RAD tags, plate barcodes, and well barcodes.

A schematic overview of Rapture.

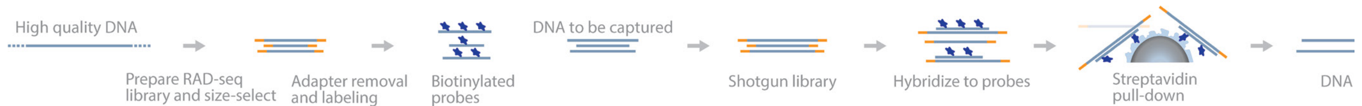


Table 16. Advantages and Disadvantages of Rapture

Advantages	Disadvantages
• Massively parallel, targeted DNA sequencing for SNP identification. • Improved number of mapped fragments and locus coverage compared to RAD-Seq.	• RAD tags needs to be designated prior to the experiment. • Additional cost to synthesize baits. • Only a small number of loci are interrogated. • Biased toward sequences closer to the restriction cut site.

Reviews

Andrews KR, Good JM, Miller MR, Luikart G and Hohenlohe PA Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat Rev Genet*. 2016;

References

Ali OA, O'Rourke SM, Amish SJ, et al. RAD Capture (Rapture): Flexible and Efficient Sequence-Based Genotyping. *Genetics*. 2016;202:389-400.

This publication describes the original Rapture protocol. To test the method, the researchers analyzed fin clips of 96 rainbow trout individuals. The average coverage at the captured loci was 16x compared, to 0.4x for RAD. In a principal-component analysis, the first principal component separated 2 distinct groups corresponding to individuals born in Bear Creek and the spring-fed spawning locations (Thousands Springs and Spring Creek). The third component separated individuals from Thousand Springs and Spring Creek. These results represent a remarkably fine scale of spatial analysis.

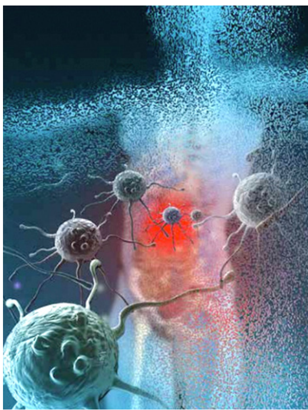
Illumina technology: HiSeq 2500 System

Associated Kits

TruSeq DNA PCR-Free Library Prep Kit

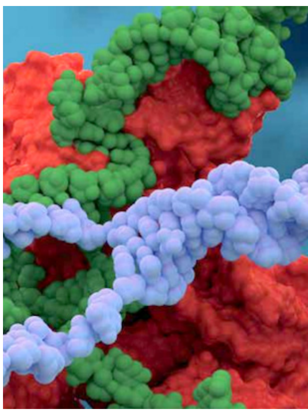
For Research Use Only. Not for use in diagnostic procedures.

Scientific Publication Reviews can be accessed at
www.illumina.com/pubreviews



Microbes and Metagenomics Research Review

Metagenomics, one of the fastest-growing scientific disciplines, is the study of DNA from microorganisms that cannot be cultured in the lab. Recent improvements to sequencing technology have generated nearly complete genome assemblies from individual microbes obtained directly from environmental or clinical research samples, without requiring special cultivation methods. This surge in metagenomics sequence information has inspired new appreciation for microbial populations and their impact on the environment and human health.



Gene Editing Research Review

CRISPR-Cas9 is a recently developed genome editing technique that allows scientists to perform precise genomic manipulation quickly and conveniently. This technology has a vast spectrum of applications. As with any molecular biology technique, it is crucial that the obtained results have high levels of specificity. This review highlights recent publications that demonstrate the use of genomic technologies and high-throughput sequencing in CRISPR-Cas9 experiments for checking specificity and off-target effects across the genome.

This Scientific Publication Review is brought to you by Illumina, Inc.
Illumina • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

FOR RESEARCH USE ONLY. Not For Use In Diagnostic Procedures.

© 2019 Illumina, Inc. All rights reserved. All trademarks are the property of Illumina, Inc. or their respective owners.
For specific trademark information, see www.illumina.com/company/legal.html
PUB 1370-2019-004 Current as of 12 June 2019

illumina®