Usos de las Diferentes Técnicas de Secuenciación

Aplicaciones de la Ingeniería Genética

Rafael Navajas Pérez

Enero de 2025

Máster en Genética y Evolución
(Especialidad Agroalimentaria)

Review Article | Published: 11 October 2017

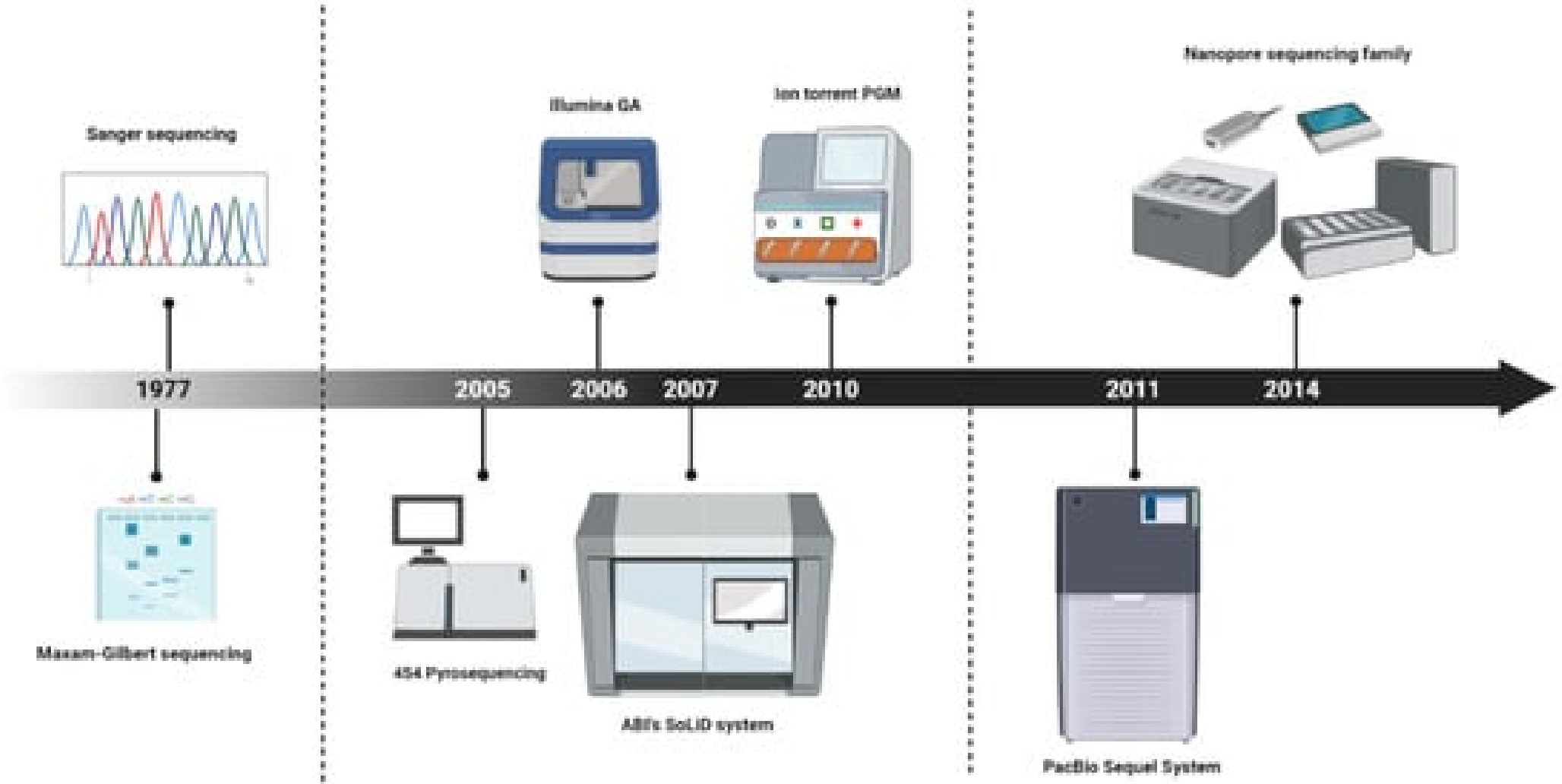
DNA sequencing at 40: past, present and future

Jay Shendure[1,2], Shankar Balasubramanian[3,4], George M. Church[5], Walter Gilbert[6], Jane Rogers[7], Jeffery A. Schloss[8] & Robert H. Waterston[1]

This review commemorates the 40th anniversary of DNA sequencing, a period in which we have already witnessed multiple technological revolutions and a growth in scale from a few kilobases to the first human genome, and now to millions of human and a myriad of other genomes. DNA sequencing has been extensively and creatively repurposed, including as a 'counter' for a vast range of molecular phenomena. We predict that in the long view of history, the impact of DNA sequencing will be on a par with that of the microscope.

The first protein sequence, of insulin, was determined in the early 1950s by Sanger, who fragmented its two chains, deciphered each fragment and overlapped the fragments to yield a complete sequence. His work showed unequivocally that proteins had defined patterns of amino acid residues[2].

In the 1960s, RNA sequencing was tackled by this same general process: an RNA species was first fragmented with RNases, next the pieces were separated by chromatography and electrophoresis, then individual fragments were deciphered by sequential exonuclease digestion, and finally the sequence was deduced from the overlaps. The first RNA sequence, of alanine tRNA, required five people working three years with one gram of pure material (isolated from 140 kg of yeast) to determine 76 nucleotides[4].



Fuente: Athanasopoulou et al., 2022

BOX 1

The milestones listed below correspond to key developments in the evolution of sequencing technologies. This is a large topic, and we apologize for any omissions.

Technical milestones

1953: Sequencing of insulin protein[2]
1965: Sequencing of alanine tRNA[4]
1968: Sequencing of cohesive ends of phage lambda DNA[6]
1977: Maxam–Gilbert sequencing[9]
1977: Sanger sequencing[8]
1981: Messing's M13 phage vector[12]
1986–1987: Fluorescent detection in electrophoretic sequencing[14,15,17]
1987: Sequenase[18]
1988: Early example of sequencing by stepwise dNTP incorporation[139]
1990: Paired-end sequencing[23]
1992: Bodipy dyes[140]
1993: *In vitro* RNA colonies[37]
1996: Pyrosequencing[44]
1999: *In vitro* DNA colonies in gels[38]
2000: Massively parallel signature sequencing by ligation[47]
2003: Emulsion PCR to generate *in vitro* DNA colonies on beads[42]
2003: Single-molecule massively parallel sequencing-by-synthesis[33,34]
2003: Zero-mode waveguides for single-molecule analysis[57]
2003: Sequencing by synthesis of *in vitro* DNA colonies in gels[49]
2005: Four-colour reversible terminators[51–53]
2005: Sequencing by ligation of *in vitro* DNA colonies on beads[41]
2007: Large-scale targeted sequence capture[93–96]
2010: Direct detection of DNA methylation during single-molecule sequencing[65]
2010: Single-base resolution electron tunnelling through a solid-state detector[141]
2011: Semiconductor sequencing by proton detection[142]
2012: Reduction to practice of nanopore sequencing[143,144]
2012: Single-stranded library preparation method for ancient DNA[145]

SECUENCIACIÓN POR TERMINADORES (Sanger)

Guidelines for Sanger sequencing and molecular assay monitoring

Beate M. Crossley,[1] Jianfa Bai, Amy Glaser, Roger Maes, Elizabeth Porter, Mary Lea Killian, Travis Clement, Kathy Toohey-Kurth

Journal of Veterinary Diagnostic Investigation
2020, Vol. 32(6) 767–775

DOI: 10.1177/1040638720905833
jvdi.sagepub.com

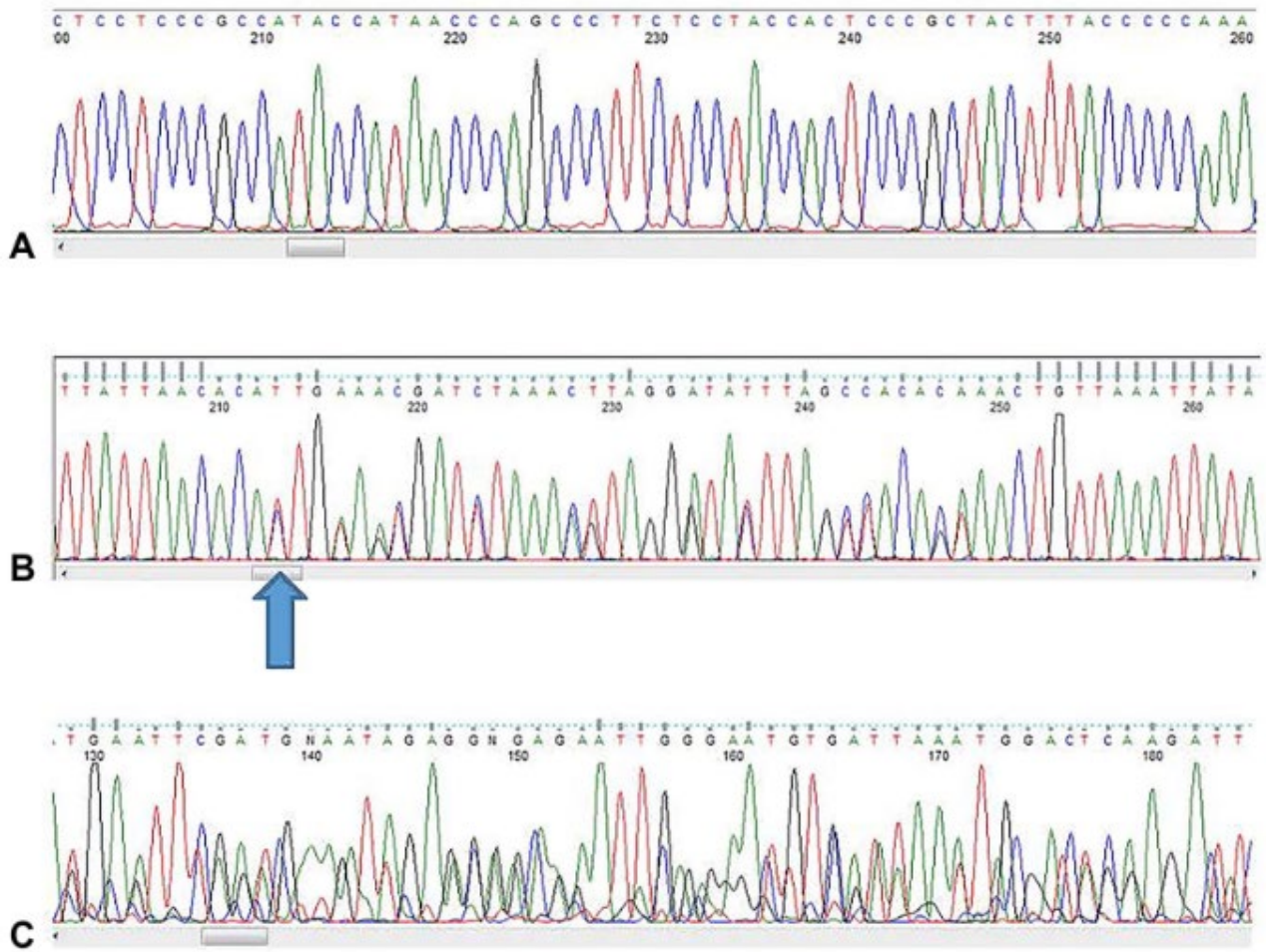


Figure 1. Examples of sequences. **A.** A good nucleic acid sequence (porcine circovirus 2). **B.** Dual infections are indicated by overlapping peaks, example pointed out by arrow (infectious bronchitis virus). **C.** High background noise (avian influenza virus), using FinchTV software v.1.4.0.

Secuenciación por Terminadores (Sanger)

SEQUENCING

Secuenciación Sanger: Quality Score (Q)

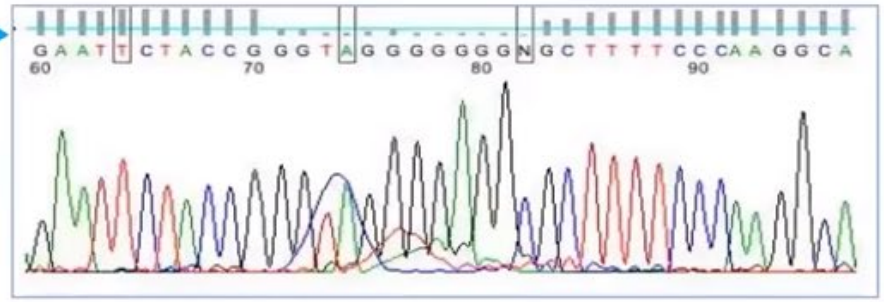
Quality Score (Q), probabilidad de que una base sea "llamada" correctamente.

Coeficiente de calidad de Phred:
Q = –10 log10 P
(P) probabilidad de que la base sea incorrecta

Quality scores and estimated base calling accuracy

Phred Quality Score	Probability of Incorrect Base Call	Base Call Accuracy
10	1 in 10	90%
20	**1 in 100**	**99%**
30*	1 in 1,000	99.9%
40	1 in 10,000	99.99%
50	1 in 100,000	99.999%

Phred score 20



Secuenciación por Terminadores (Sanger)

NGS

NGS basic concepts

Sensitivity: probability of detecting a mutation at a given allele frequency or abundance level of a tumor clone.

Fold Coverage	P_0	Percent Not Sequenced	Percent Sequenced
0.25	$e^{-0.25} = 0.78$	78	22
0.5	$e^{-0.5} = 0.61$	61	39
0.75	$e^{-0.75} = 0.47$	47	53
1	$e^{-1} = 0.37$	37	63
2	$e^{-2} = 0.135$	13.5	87.5
3	$e^{-3} = 0.05$	5	95
4	$e^{-4} = 0.018$	1.8	98.2
5	$e^{-5} = 0.0067$	0.6	99.4
6	$e^{-6} = 0.0025$	0.25	99.75
7	$e^{-7} = 0.0009$	0.09	99.91
8	$e^{-8} = 0.0003$	0.03	99.97
9	$e^{-9} = 0.0001$	0.01	99.99
10	$e^{-10} = 0.000045$	0.005	99.995

Probabilidad de que un nucleótido cualquiera no esté cubierto

En función de la sensibilidad deseada así seleccionaremos la profundidad de cobertura:

Se estima una cobertura mínima de 5x a 10x para que se pueda afirmar que el nucleótido está presente.

- Cobertura mínima: 5-10x
- Detección de heterozigotos: 30-40x
- Variantes presentes en el 5%: mínimo 500x

- Para aceptar una variante detectada debemos verla en al menos 10 lecturas

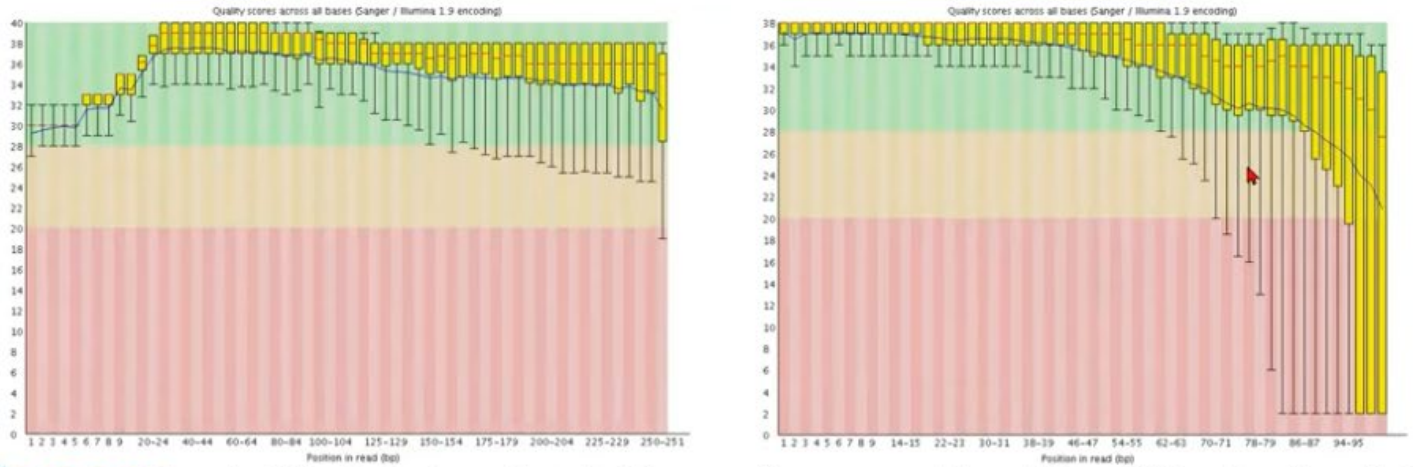
Secuenciación por Terminadores (Sanger)

NGS

NGS basic concepts

Next Generation Sequencing (NGS) is a general term which describes several different sequencing technologies allowing us to identify a wide range of genetic alterations.

Key concepts in NGS



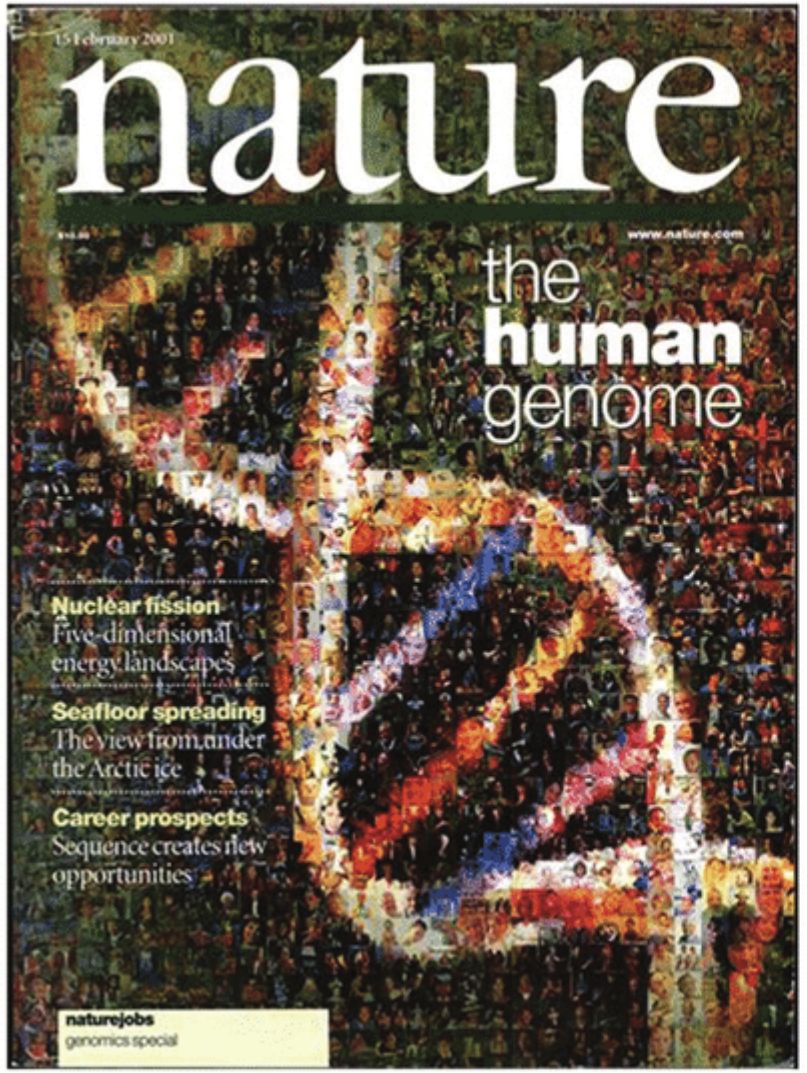
Quality score: probability that a base is called incorrectly, assessed by a "Phred"-like algorithm → Q=30

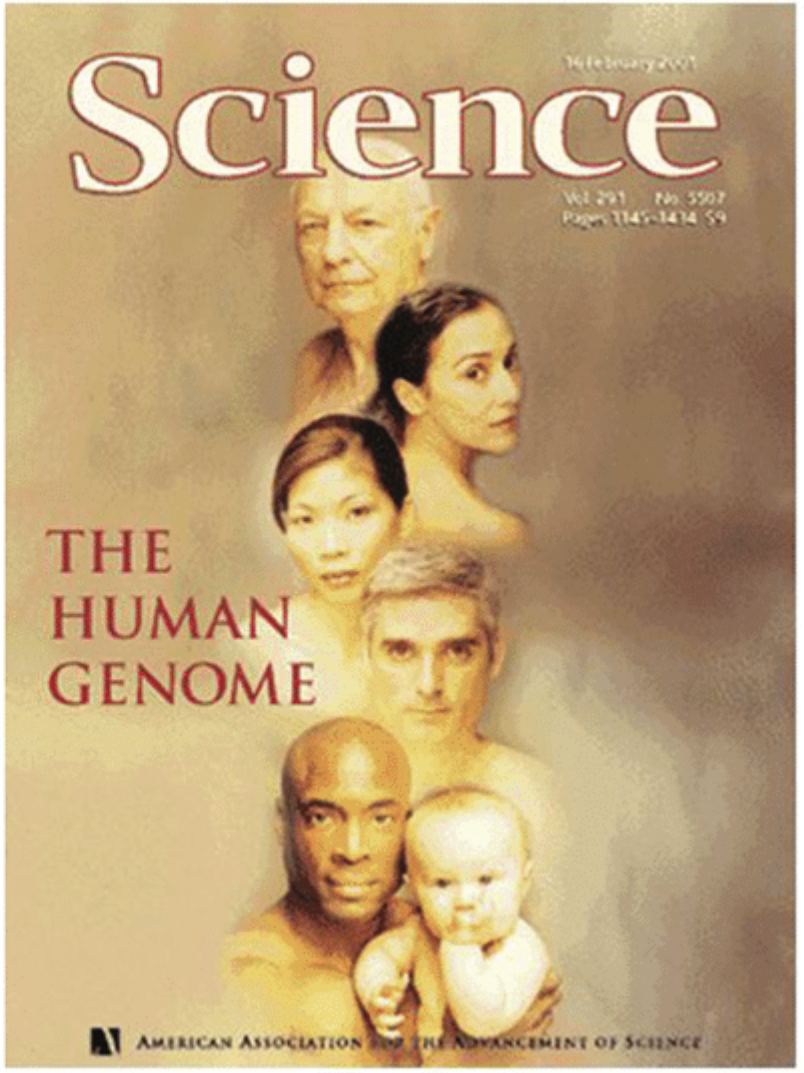
Shotgun jerárquico:

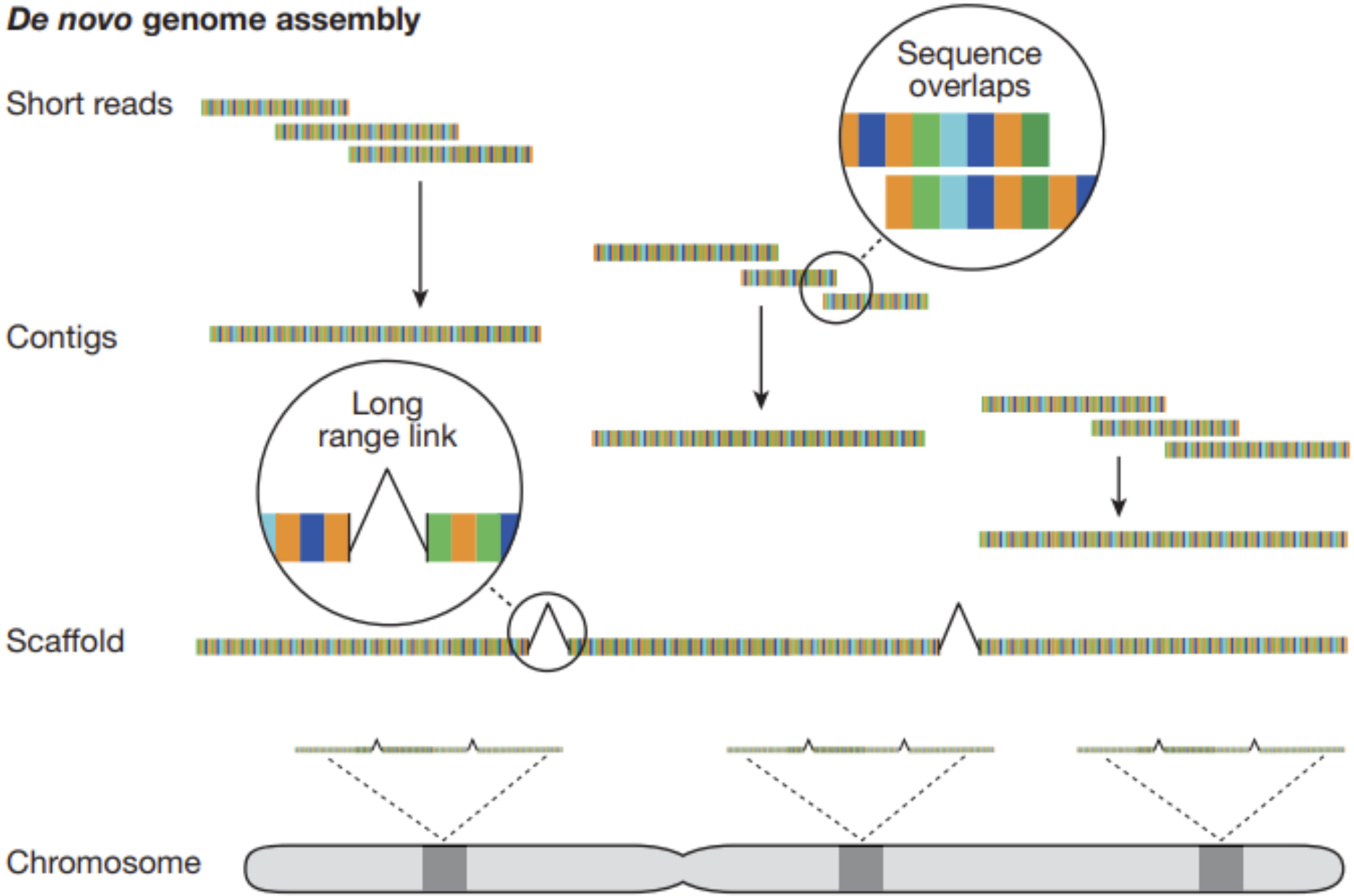
- El **Proyecto Genoma Humano** supone la punta de lanza de esta tecnología.
- Era necesario clonar grandes fragmentos del genoma humano en **cromosomas artificiales de bacterias (BACs)**.
- Cada BAC era, a su vez, **fragmentado, seleccionado** (tamaño, marcadores) **y clonado**.
- A partir de cada clon, se obtenía el **ADN que iba a ser secuenciado**.

Shotgun genoma completo:

- Basado en el **solapamiento** de gran número de fragmentos.
- **Sin mapeo físico**.







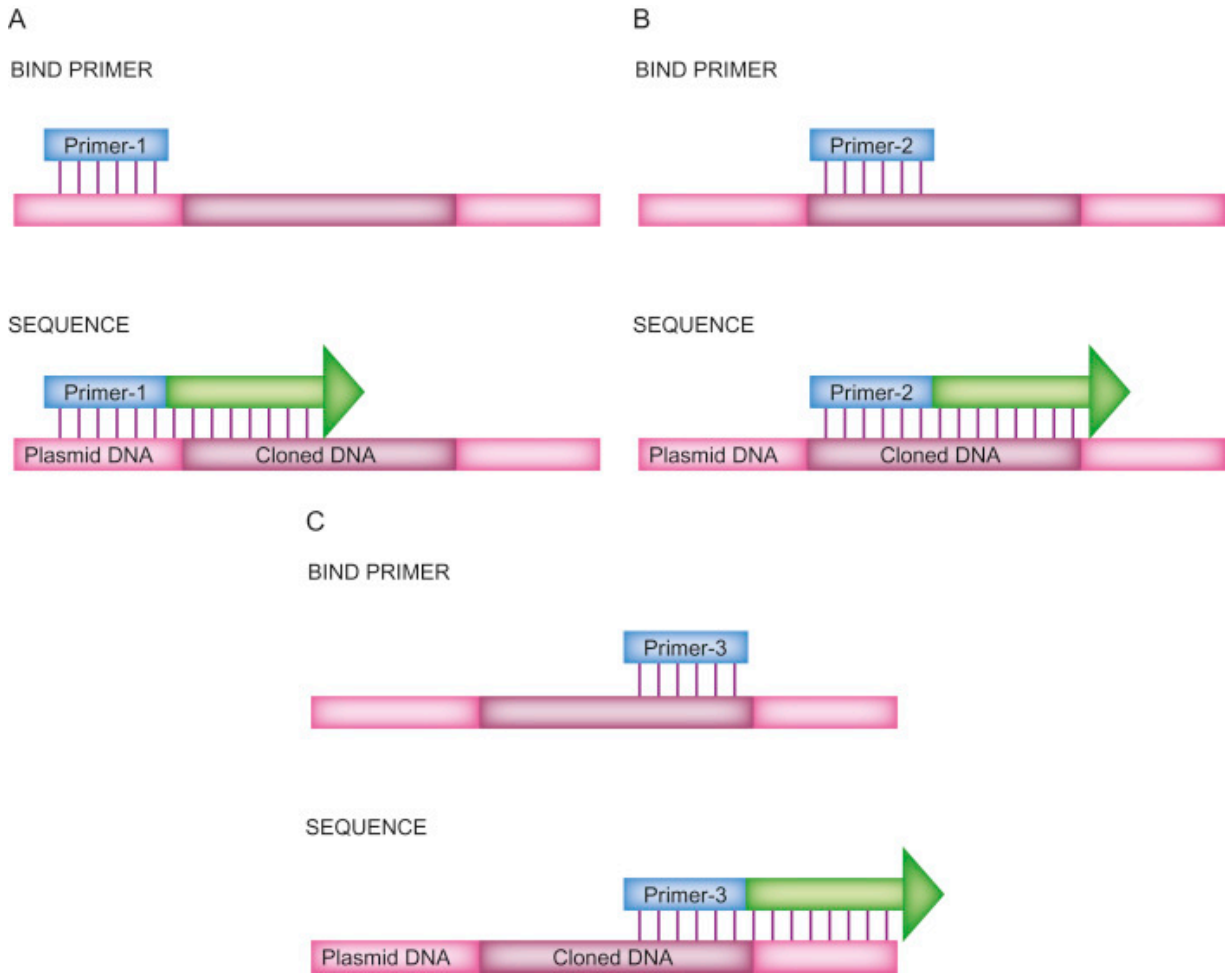
Fuente: Shendure et al., 2017

Uno de los problemas fundamentales de esta técnica lo constituyen las **discontinuidades**.

Las secuencias pareadas se agrupaban en **contigs**. Aquellos contigs solapantes constituían **scaffolds** de mayor tamaño.

Los huecos existentes entre scaffolds tenían que ser 'cerrados' mediante **secuenciación directa y *chromosome walking***.

Proceso llevado a cabo de forma manual por científicos **'finishers'.**



the plant journal

SEB Society for Experimental Biology

The Plant Journal (2011) **67**, 305–317 doi: 10.1111/j.1365-313X.2011.04590.x

Micro-collinearity and genome evolution in the vicinity of an ethylene receptor gene of cultivated diploid and allotetraploid coffee species (*Coffea*)

Qingyi Yu[1,2,*,†], Romain Guyot[3,†], Alexandre de Kochko[3], Anne Byers[1], Rafael Navajas-Pérez[4], Brennick J. Langston[2], Christine Dubreuil-Tranchant[3], Andrew H. Paterson[4], Valérie Poncet[3], Chifumi Nagai[1] and Ray Ming[1,5]

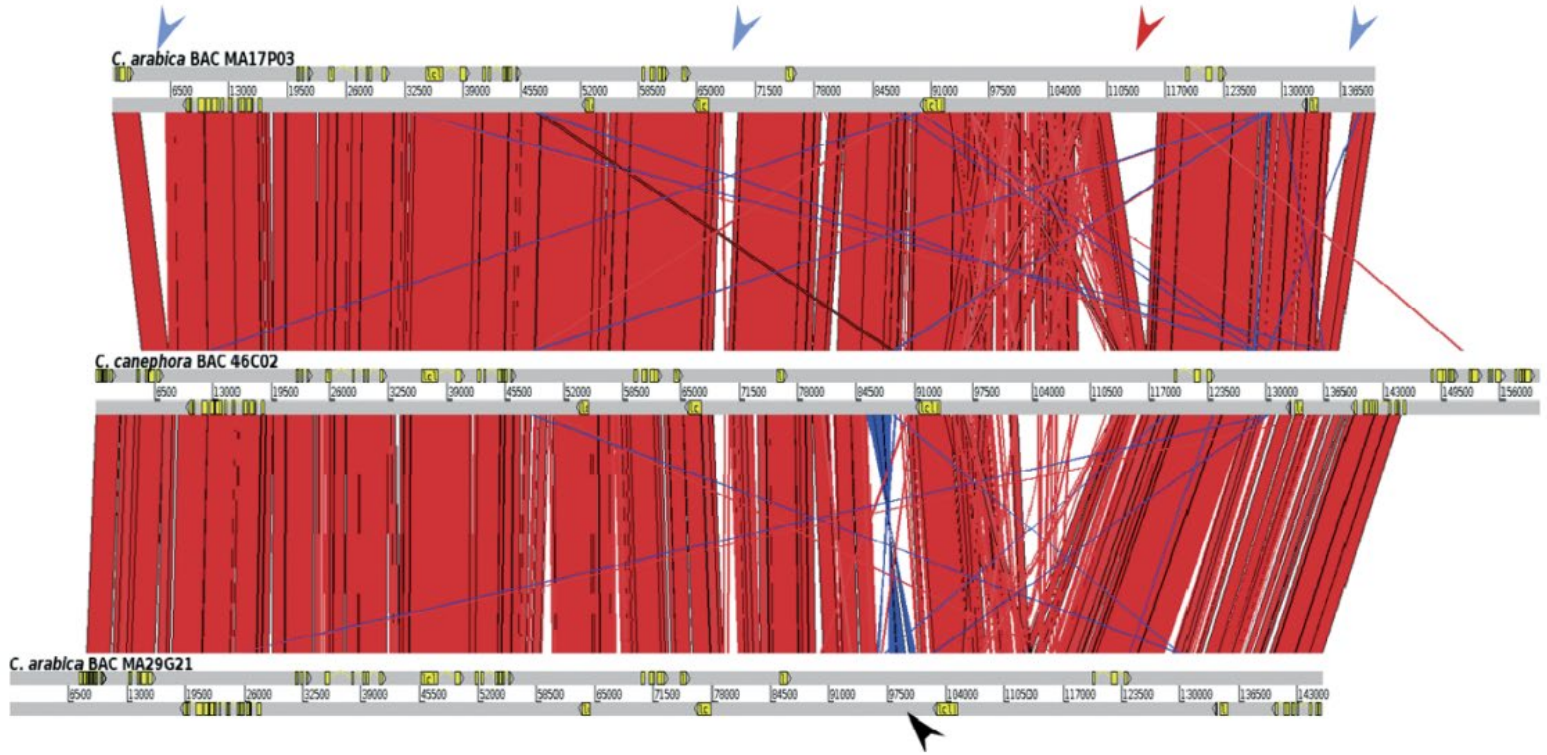
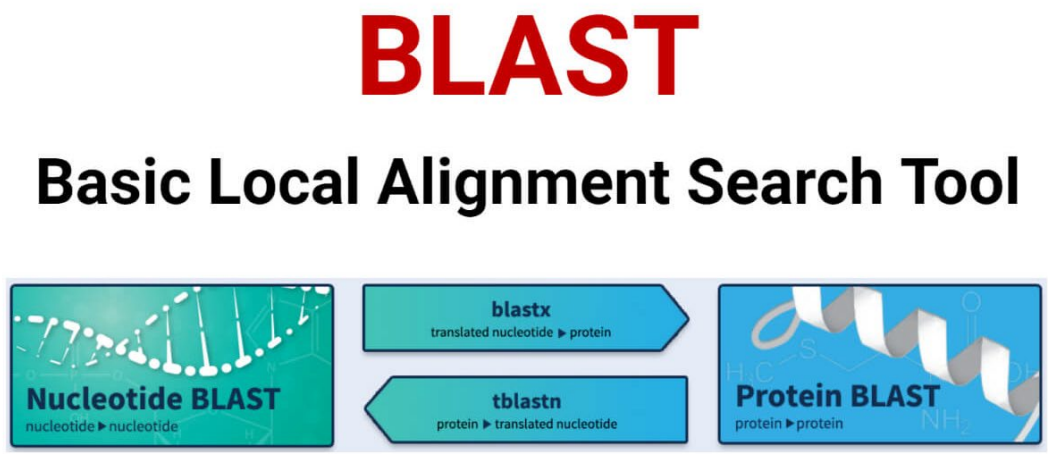


Figure 1. Sequence comparison between *Coffea canephora* (46C02) and *C. arabica* (MA29G21 and MA17P03).
The red lines link conserved regions and blank regions symbolize disrupted conservation. The blue and red arrows represent insertion of transposons and retrotransposons, respectively, in the MA17P03 BAC sequence. The black arrow indicates a chromosomal inversion. Comparisons were performed using the BLASTZ algorithm (Schwartz *et al.*, 2003) and visualized using the Artemis Comparison Tool (Carver *et al.*, 2005).

Secuenciación por Terminadores (Sanger)

El **volumen de datos generados** crece exponencialmente.
Nace la necesidad de crear **repositorios** para centralizar la información (GenBank) y herramientas que sean capaces de interrogar a las bases de datos (BLAST).
En 1982, en torno a **medio millón de bases** se habían depositado en GenBank; En 1986, la cifra casi llegaba a los **10 millones de bases** (https://www.ncbi.nlm.nih.gov/genbank/statistics/).



Fuente: Tamang, 2023

NGS-SBS
(Next Generation Sequencing-Sequencing By Synthesis)

SOLEXA (Ilumina): usos

	Secuenciación del genoma humano completo	Genomas grandes y complejos	Genomas pequeños
Aplicaciones de secuenciación	• Investigación genómica del cáncer • Detección de variantes • Estudios de riesgos genéticos • Genética de poblaciones	• Agrigenómica (maíz, trigo, res vacuna, etc.) • Organismos modelo (mosca de la fruta, ratón, pez cebra, etc.) • Investigación de plantas y animales	• Microbioma humano • Microbiología/metagenómica • Investigación de salud pública • Secuenciación de amplicones
Ejemplos de secuenciación	• Genoma humano (3,2 Gb), 30×, S2 kit, sistema NovaSeq™, 8 muestras/celda de flujo • Genoma humano (3,2 Gb), >30×, v2.5 sistema HiSeq X, 8 muestras/celda de flujo	• Genoma de la mosca de la fruta (175 Mb), 30×, v2 kit, sistema NextSeq™ 550, 22 muestras/celda de flujo • Genoma del ratón (2,7 Gb), 30×, v1 kit sistema HiSeq 4000 System, 8 muestras/celda de flujo	• Genoma *E. coli* (4,6 Mb), 30×, sistema MiniSeq™, 50 muestras/celda de flujo • Plásmidos o amplicones (650 kb), 1000× MiSeq System, 11 muestras/celda de flujo

Figura 6: Amplia gama de aplicaciones con Illumina DNA Prep. Desde la secuenciación del genoma humano completo y genomas grandes y complejos, hasta la de pequeños genomas microbianos, Illumina DNA Prep ofrece flexibilidad en todos los experimentos.

Pirosecuenciación 454 (Life Sciences): usos

NOTAS DE PRENSA

11 de junio de 2008

74

Genómica General

El Secuenciador 454 de Roche estudiará el genoma del melón en España

Un nuevo sistema de secuenciación: el Genome Sequencer FLX de Roche apoyará estas tareas de investigación

SNPs

SECUENCIACIÓN DE TERCERA GENERACIÓN

Received: 10 January 2019 | Accepted: 22 April 2019

DOI: 10.1111/2041-210X.13201

PRACTICAL TOOLS

Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

A rapid environmental DNA method for detecting white sharks in the open ocean

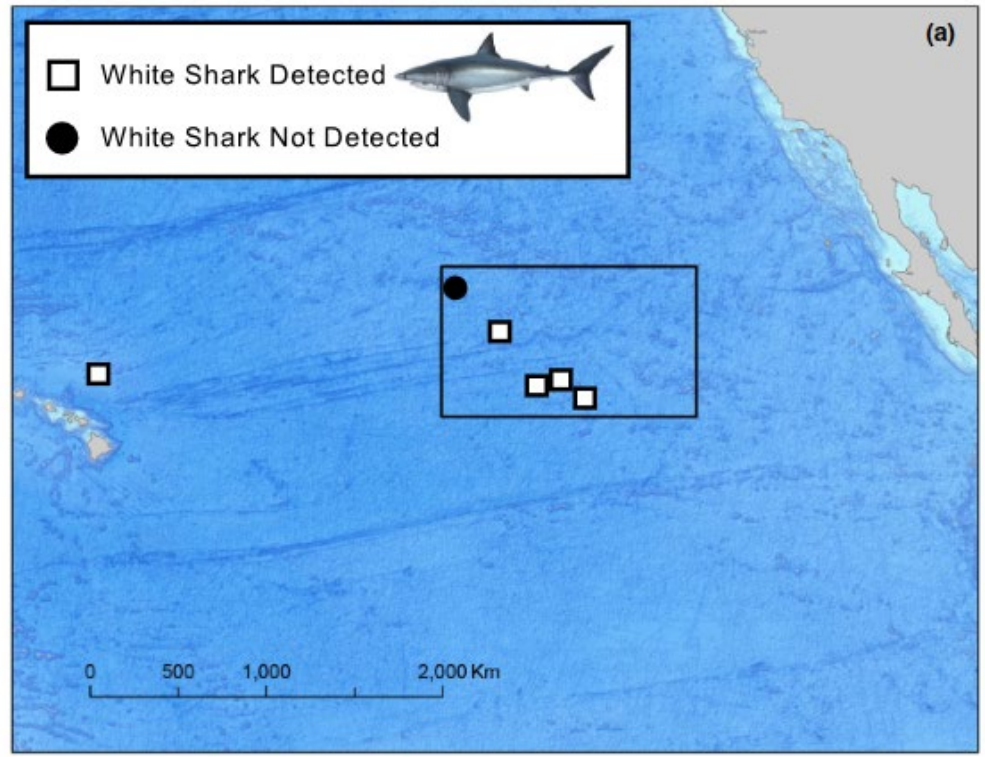
Nathan K. Truelove[1] | Elizabeth A. Andruszkiewicz[2] | Barbara A. Block[1]

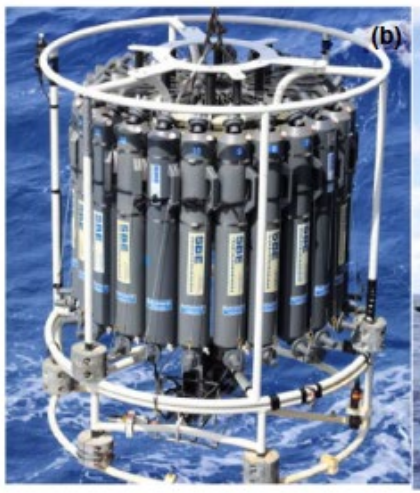
2.4 | PCR amplification

Vertebrate eDNA was amplified from eDNA extracts using the MiFish universal fish primers (Miya et al., 2015) targeting the 12S rRNA gene in 20 µl reactions. The primer sequences utilized

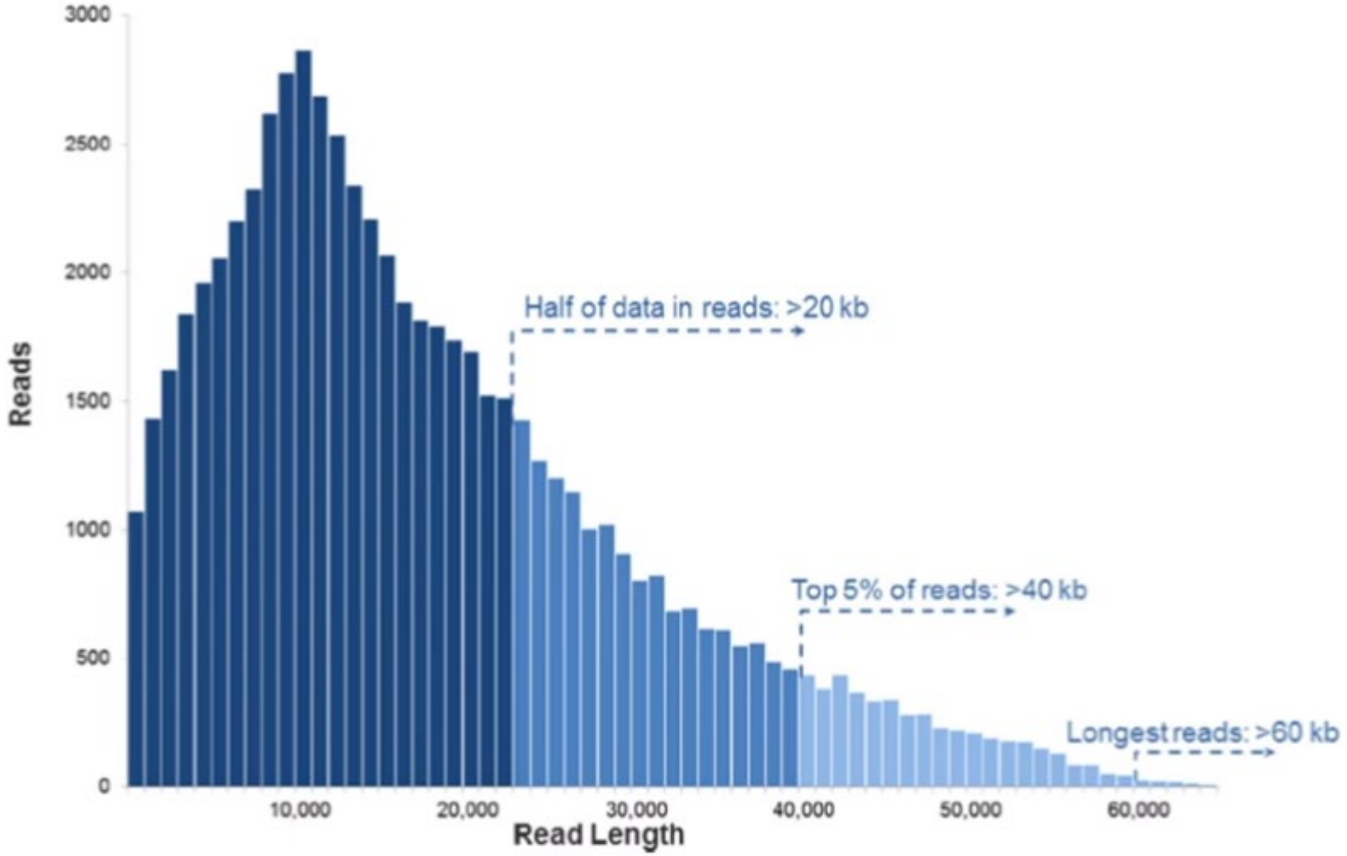
2.6 | MinIon sequencing

A total of 10 libraries were successfully prepared using Oxford Nanopore's 1D Amiplicon by ligation protocol (SQK-LSK108, Oxford Nanopore, UK). Six libraries contained a pooled depth sample consisting of two to six depths and four libraries consisted of a single depth (Table S1). Due to limited Internet connection, we used MinKnow offline v18.1.6.0 without live base calling to sequence. The "Platform QC" script was run on each flow cell (R9.4 version). The MinIon sequencer was run for ~3.5 hr for each run.





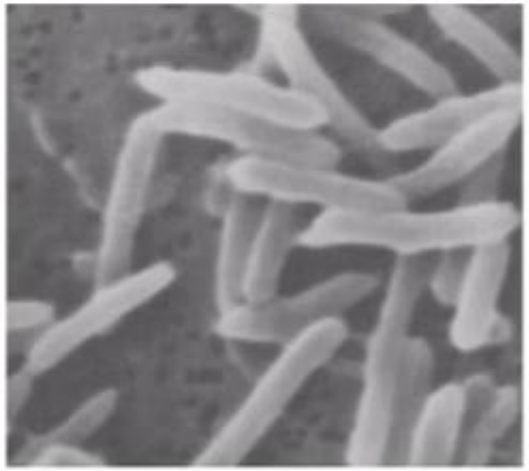
PacBio (Pacific Biotechnology)



Genomics Proteomics Bioinformatics, 13: 278-289

PacBio (Pacific Biotechnology)

Clostridium autoethanogenum



Produce etanol a partir de CO
Genoma de 4.35 Mb

Table 2 Assembly statistics for strain DSM 10061

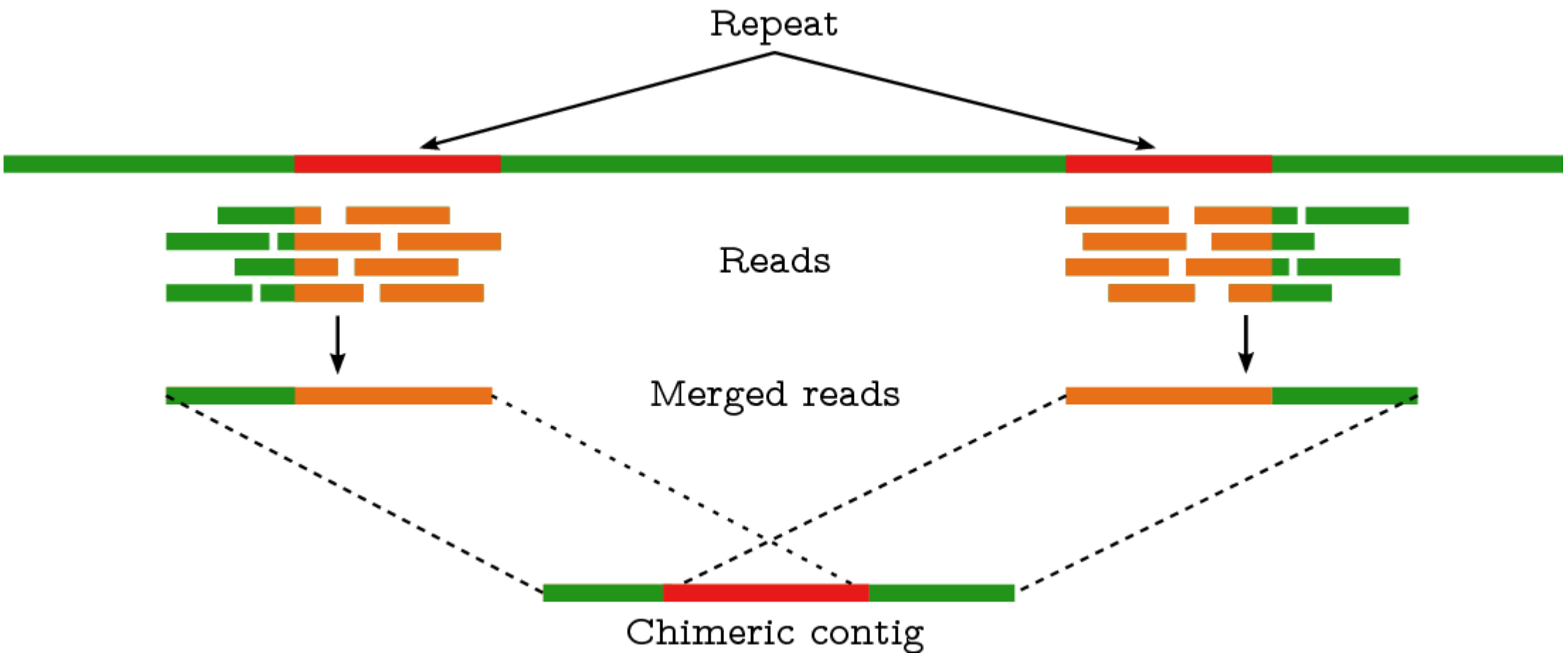
	Contigs (number)	Largest contig (bp)	Contig N50 (bp)	Genome size (Mb)	Scaffolds	Largest scaffold (bp)
454/Ion Torrent *	100	436,795	115,901	4.32		
Illumina only	57	460,940	255,482	4.3	53	769,812
454 only	32	134,546	330,116	4.3	13	1,137,876
Illumina/ 454 Hybrid	22	1,137,625	687,076	4.3	13	1,137,625
PacBio	1	4,352,205	4,352,267	4.3	1	4,352,267

Biotechnology for Biofuels, 7: 40

Corpus ID: 54888976

Evaluation of DNA scaffolding techniques using PacBio long reads

K. Patsis • Published 17 September 2014 • Biology, Computer Science, Environmental Science



Phase 1 (contig assembly)

(a) Sequence genome to produce short reads

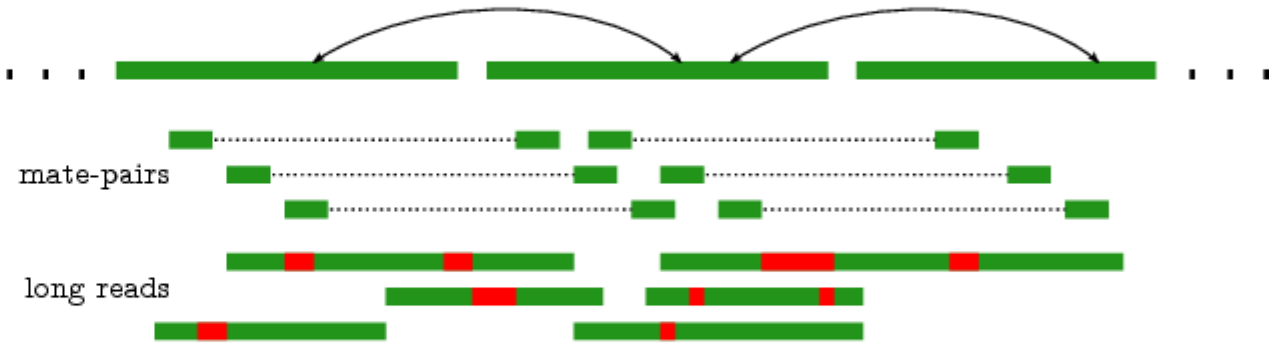
(b) Find overlaps between reads

GATGGACAACCGAACGGTCA
GAACGCTCATATAGTCAAATGG

(c) Assemble unambiguous overlaps into contigs

Phase 2 (scaffolding)

(d) Map mate-pairs or long reads to contigs and identify links



(e) Connect linked contigs into scaffolds

Genomics Proteomics Bioinformatics 13 (2015) 278–289

PacBio Sequencing and Its Applications

Anthony Rhoads [1,a], **Kin Fai Au** [1,2,*,b]

Closing gaps in draft genomes can also be accomplished efficiently via PacBio sequencing of PCR products. This approach is more cost-effective than Sanger sequencing and is able to close gaps greater than 2.5 kb in a single round of reactions

STRs are associated with many genetic disorders and are difficult to detect with SGS technologies. One such gene is the human fragile X mental retardation 1 (*FMR1*) gene. *FMR1* contains a $(CGG)_n$ repeat that is responsible for heritable disorders including fragile X syndrome, fragile X-associated tremor/ataxia syndrome, adult-onset neurodegenerative disorder, premature ovarian insufficiency, learning disabilities, autism spectrum disorders, attention deficit hyperactivity disorder, and seizures [41]. There are normally 7–60 (CGG) repeats, while the permutation range is 60–230 repeats, and the full mutation range is over 230 repeats [42]. Loomis et al. generated PacBio long reads for expanded CGG-repeat *FMR1* alleles in full mutation range [41]. They demonstrated that PacBio sequencing was not adversely affected by expansions exceeding 750 repeats, suggesting that productive sequencing is limited only by factors governing the productive lifetime of the polymerase and the desired number of subreads within an individual CCS read. PacBio targeted sequencing has also been used to resolve the genomic gap in *MUC5AC* [43], which encodes a large, secreted mucin that is important in cystic fibrosis, lung cancer, and respiratory diseases [44]. By

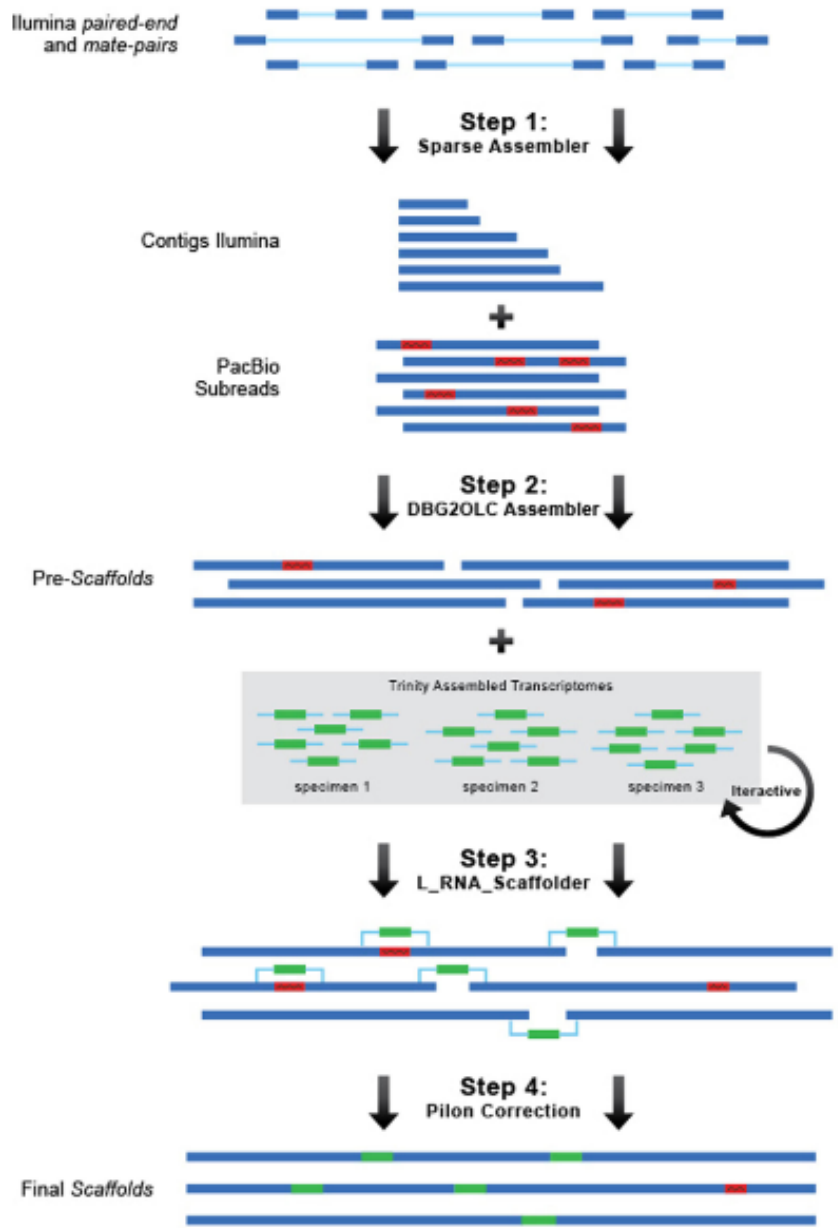
GigaScience, 7, 2018, 1–10

doi: 10.1093/gigascience/gix128
Advance Access Publication Date: 15 December 2017
Data Note

DATA NOTE

A hybrid-hierarchical genome assembly strategy to sequence the invasive golden mussel, *Limnoperna fortunei*

Marcela Uliano-Silva[1,2,3,*], Francesco Dondero[4], Thomas Dan Otto[5,6], Igor Costa[7], Nicholas Costa Barroso Lima[7,8], Juliana Alves Americo[1], Camila Junqueira Mazzoni[2,3], Francisco Prosdocimi[7] and Mauro de Freitas Rebelo[1,*]



nature plants

ARTICLES

https://doi.org/10.1038/s41477-021-01031-8

OPEN

Representation and participation across 20 years of plant genome sequencing

Rose A. Marks[1,2,3]✉, Scott Hotaling[4], Paul B. Frandsen[5,6] and Robert VanBuren[1,2]

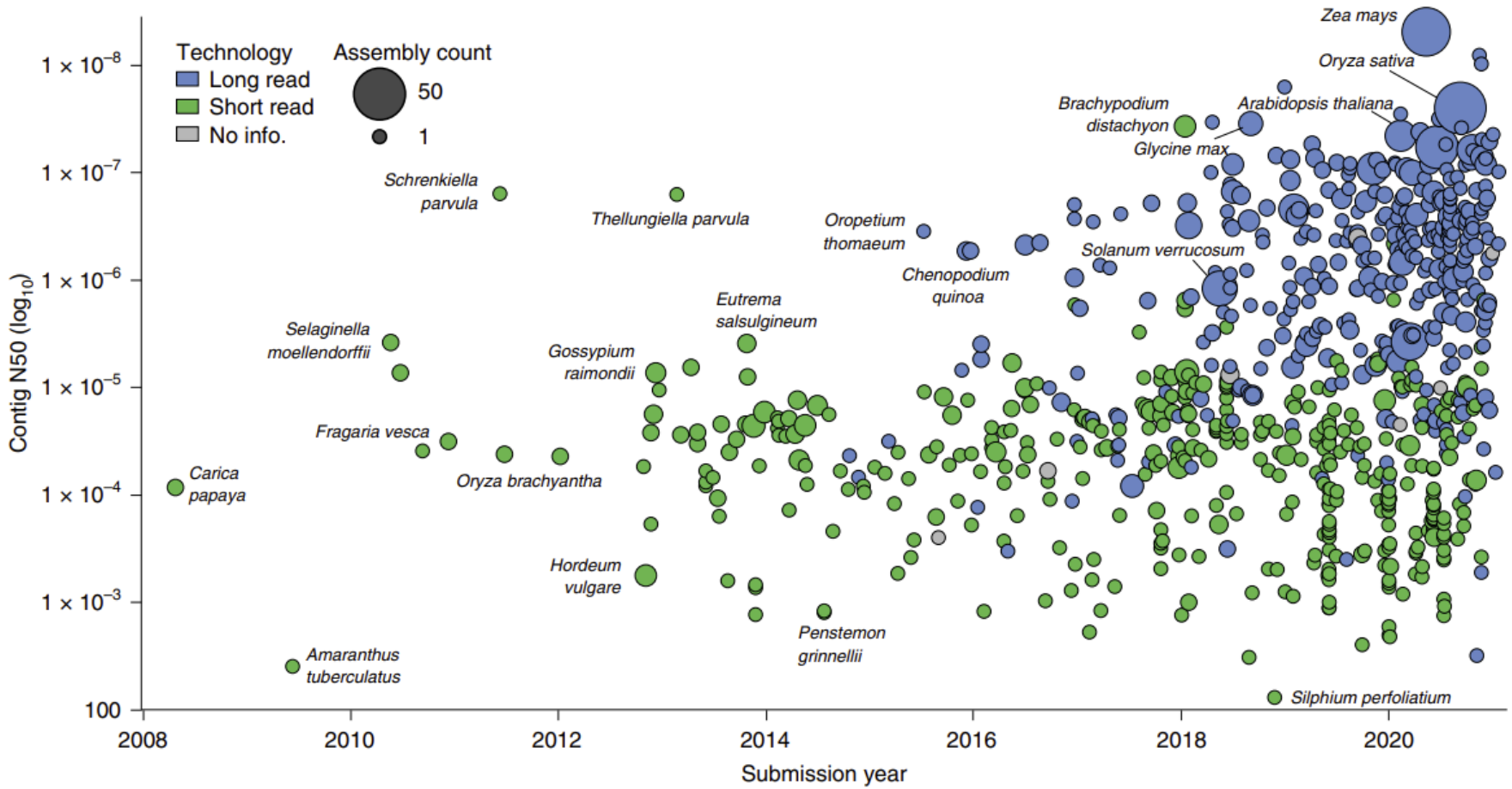


Fig. 1 | Changes in land plant genome assembly quality and availability over time. Assembly contiguity by submission date for 798 land plant species with publicly available genome assemblies. Points are coloured by the type of sequencing technology used and scaled by the number of assemblies available for that species. There is an improvement in contiguity associated with the advent of long-read sequencing technology, and a noticeable increase in the number of genome assemblies generated annually. All assemblies generated before 2008 have since been updated and are therefore not included.

	Scientific name	Common name	Year	Type	Division or monocot/ dicot	Chr (#)	Size	Assembled	Assem	Gene (#)	Repeat	scaffold N50	contig N50	Sequencer types	Journal	PMID
							Mb		%		%	kb				
1	*Arabidopsis thaliana*	arabidopsis	2000	model	dicot	5	125	115	92	25,498	14	NA	NA	Sa	Nature	11130711
2	*Oryza sativa*	rice	2002	crop	monocot	12	430	362	84	59,855	26	12	7	Sa	Science	11935017
3	*Oryza sativa*	rice	2002	crop	monocot	12	420	389	93	61,668	NA	NA	NA	Sa	Science	11935018
4	*Oryza sativa*	rice	2005	crop	monocot	12	389	371	95	37,544	26	NA	NA	Sa	Nature	16100779
5	*Populus trichocarpa*	black cottonwood	2006	crop	dicot	19	485	410	84	45,555	NA	3100	126	Sa	Science	16973872
6	*Vitis vinifera*	grape	2007	crop	dicot	19	475	487	103	30,434	41	2065	66	Sa	Nature	17721507
7	*Physcomitrella patens*	moss	2008	model	bryophyta	27	510	480	94	35,938	16	1320	292	Sa	Science	18079367
8	*Vitis vinifera*	grape	2007	crop	dicot	19	505	477	95	29,585	27	1330	18	Sa,4	PlosOne	18094749
9	*Carica papaya*	papaya	2008	crop	dicot	9	372	370	99	28,629	43	1000	11	Sa	Nature	18432245
10	*Lotus japonicus*	lotus	2008	model	dicot	6	472	315	67	30,799	56	NA	NA	Sa	DNA Research	18511435
11	*Sorghum bicolor*	sorghum	2009	crop	monocot	10	818	739	90	34,496	62	62,400	195	Sa	Nature	19189423
12	*Cucumis sativus*	cucumber	2009	crop	dicot	7	367	244	66	26,682	24	1140	20	Sa,I	Nature Genetics	19881527
13	*Zea mayes*	maize	2009	crop	monocot	10	2300	2048	89	32,540	85	76	40	Sa	Science	19965430
14	*Glycine max*	soybean	2010	crop	dicot	20	1115	973	87	46,430	57	47,800	189	Sa	Nature	20075913
15	*Brachypodium distachyon*	brachypodium	2010	model	monocot	5	272	272	100	25,532	21	59,300	348	Sa	Nature	20148030
16	*Ricinus communis*	castor bean	2010	crop	dicot	10	320	326	102	31,237	50	561	21	Sa	Nature Biotechnology	20729833
17	*Malus x domestica*	apple	2010	crop	dicot	17	742	604	81	57,386	67	1542	13	Sa,4	Nature Genetics	20802477
18	*Jatropha curcas*	jatropha	2010	crop	dicot	NA	380	286	75	40,929	37	NA	4	Sa,	DNA Research	21149391
19	*Theobroma cacao*	cocoa	2011	crop	dicot	10	430	327	76	28,798	24	473	20	Sa,4,I	Nature Genetics	21186351
20	*Fragaria vesca*	strawberry	2011	crop	dicot	7	240	210	87	34,809	23	1361	NA	4,S,I	Nature Genetics	21186353
21	*Arabidopsis lyrata*	lyrata	2011	model	dicot	8	207	207	100	32,670	30	24,500	227	Sa	Nature Genetics	21478890
22	*Selaginella moellendorffii*	spikemoss	2011	non-model	lycopod	NA	110	213	193	22,285	38	1700	120	Sa	Science	21551031
23	*Phoenix dactylifera*	date palm	2011	crop	monocot	18	658	381	58	28,890	40	30	6	I	Nature Biotechnology	21623354
24	*Solanum tuberosum*	potato	2011	crop	dicot	12	844	727	86	39,031	62	1318	31	Sa,4,I	Nature	21743474
25	*Thellungiella parvula*	thellungiella	2011	model	dicot	7	140	137	98	30,419	8	5290	NA	4,I	Nature Genetics	21822265
26	*Cucumis sativus*	cucumber	2011	crop	dicot	7	367	323	88	26,587	NA	319	323	Sa,4	PlosOne	21829493
27	*Brassica rapa*	chinese	2011	crop	dicot	10	485	284	59	41,174	40	1971	27	I	Nature Genetics	21873998

	Scientific name	Common name	Year	Type	Division or monocot/ dicot	Chr (#)	Size	Assembled	Assem	Gene (#)	Repeat	scaffold N50	contig N50	Sequencer types	Journal	PMID
							Mb		%		%	kb				
28	*Cannabis sativa*	hemp	2011	crop	dicot	?	820	787	96	30,074	NA	16	2	4,I	Genome Biology	22014239
29	*Cajanus cajan*	pigeon pea	2011	crop	dicot	11	833	605	72	48,680	52	516	22	Sa,I	Nature Biotechnology	22057054
30	*Mediucago truncatula*	medicago	2011	model	dicot	8	454	262	58	62,388	31	1270	NA	Sa,4,I	Nature	22089132
31	*Setaria italica*	setaria	2012	model	monocot	9	490	423	86	38,801	46	1007	25	I	Nature Biotechnology	22580950
32	*Setaria italica*	setaria	2012	model	monocot	9	510	397	80	35,471	40	47,300	126	Sa	Nature Biotechnology	22580951
33	*Solanum lycopersicum*	tomato	2012	crop	dicot	12	900	760	84	34,727	63	16,467	87	Sa,4,S,I	Nature	22660326
34	*Cucumis melo*	melon	2012	crop	dicot	12	450	375	83	27,427	NA	4680	18	Sa,4,I	PNAS	22753475
35	*Linum usitatissimum*	flax	2012	crop	dicot	15	373	318	85	43,484	24	132	20	I	Plant Journal	22757964
36	*Musa acuminata malaccensis*	banana	2012	crop	monocot	11	523	472	90	36,542	44	1311	43	Sa,4,I	Nature	22801500
37	*Gossypium raimondii*	cotton D	2012	crop	dicot	13	880	775	88	40,976	60	2284	45	I	Nature Genetics	22922876
38	*Azadirachta indica*	neem	2012	crop	dicot	NA	364	NA	NA	20,169	13	452	1	4,I	BMC Genomics	22958331
39	*Hordeum vulgare*	barely	2012	crop	monocot	7	5100	4980	98	30,400	84	NA	NA	NA	Nature	23075845
40	*Pyrus bretschneideri*	pear	2013	crop	dicot	17	527	512	97	42,812	53	541	36	I	Genome Research	23149293
41	*Citrullus lanatus*	watermelon	2012	crop	dicot	11	425	354	83	23,440	45	2380	26	I	Nature Genetics	23179023
42	*Triticum aestivum*	wheat	2012	crop	monocot	21	17,000	3800	22	94,000	80	NA	1	4	Nature	23192148
43	*Gossypium raimondii*	cotton D	2012	crop	dicot	13	880	738	84	37,505	61	18,800	136	Sa,4,I	Nature	23257886
44	*Prunus mume*	chinese plum	2012	crop	dicot	8	280	237	85	31,390	45	578	32	I	Nature Communications	23271652
45	*Cicer arietinum*	chickpea	2013	crop	dicot	8	738	532	72	28,269	49	39,990	24	Sa,I	Nature Biotechnology	23354103
46	*Hevea brasiliensis*	rubber tree	2013	crop	dicot	18	2150	1119	52	68,955	72	3	NA	4,S,I	BMC Genomics	23375136
47	*Phyllostachys heterocycla*	moso bamboo	2013	non-model	monocot	24	2075	2051	99	31,987	59	329	12	I	Nature Genetics	23435089
48	*Oryza brachyantha*	rice relative	2013	non-model	monocot	12	300	263	88	32,038	29	1013	20	I	Nature Communications	23481403
49	*Prunus persica*	peach	2013	crop	dicot	8	265	227	86	27,852	37	27,400	214	Sa	Nature Genetics	23525075
50	*Aegilops tauschii*	wheat DD	2013	crop	monocot	7	4360	4244	97	43,150	66	58	5	4,I	Nature	23535592

	Scientific name	Common name	Year	Type	Division or monocot/ dicot	Chr (#)	Size	Assembled	Assem	Gene (#)	Repeat	scaffold N50	contig N50	Sequencer types	Journal	PMID
							——Mb——		%		%	kb				
51	*Triticum urartu*	wheat AA	2013	crop	monocot	7	4940	4660	94	34,879	67	64	3	I	Nature	23535596
52	*Nelumbo nucifera*	ancient lotus	2013	non-model	dicot	8	929	804	87	26,685	57	3400	39	I	Genome Biology	23663246
53	*Utricularia gibba*	bladderwort	2013	non-model	dicot	16	77	82	106	28,500	3	95	26	4,I	Nature	23665961
54	*Picea abies*	norway spruce	2013	crop	gymnosperm	12	19,600	12,000	61	28,354	NA	NA	NA		Nature	23698360
55	*Capsella rubella*	capsella	2013	non-model	dicot	8	219	135	62	26,521	NA	15,100	134	Sa	Nature Genetics	23749190

†Abbreviations: Sa, Sanger; 4, Roche/454; S, SOLiD; I, Illumina; NA, not reported in primary publication; kb, kilobases; Mb, megabases; Chr, chromosome; PMID, PubMed ID

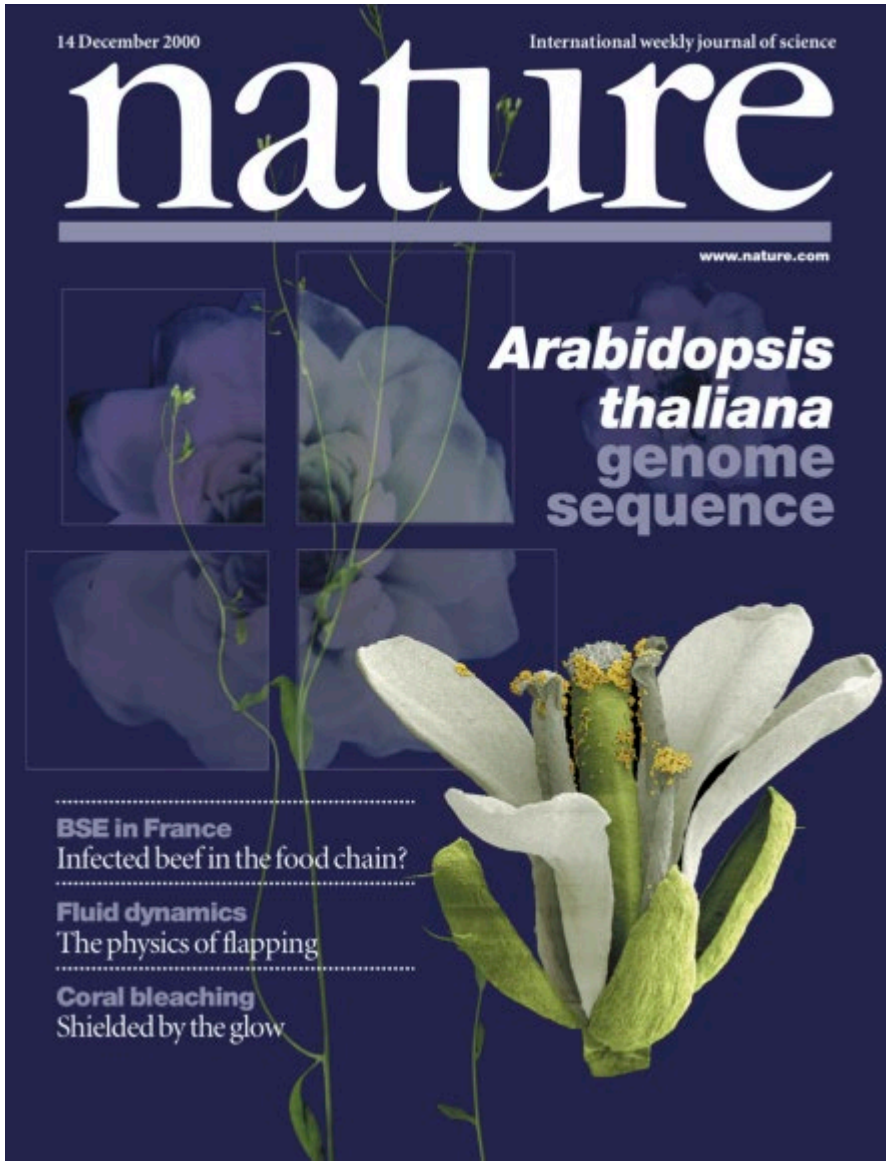
The First 50 Plant Genomes

Todd P. Michael* and Scott Jackson

Published in The Plant Genome 6
doi: 10.3835/plantgenome2013.03.0001in

5585 Guilford Rd., Madison, WI 53711 USA
An open-access publication

Hoy, disponemos de unos **800 genomas vegetales** (ensamblados a distintos niveles).


14 December 2000
International weekly journal of science
nature
www.nature.com
Arabidopsis thaliana genome sequence
BSE in France
Infected beef in the food chain?
Fluid dynamics
The physics of flapping
Coral bleaching
Shielded by the glow


5 April 2002
Science
Vol. 296 No. 5565
Pages 1–204 $9
The RICE Genome
American Association for the Advancement of Science

Supplementary Table 1. Statistics of whole genome shotgun end reads and BAC end reads.

Insert size (kb)	Vector	LUCY trimmed bases (billions)	LUCY trimmed bases after removing organellar reads (billions)	Number of reads after removing organellar reads (millions)	Sequence coverage by LUCY trimmed bases	Sequence coverage by bases with quality ≥ 20	Fraction of paired reads (%)	Fraction of assembled reads (%)
3	Plasmid	1.01	0.67	0.86	1.80 X	1.64 X	95.2	88.1
6	Plasmid	0.68	0.51	0.69	1.36 X	1.25 X	93.2	85.0
86	BAC	0.02	0.02	0.03	0.06 X	0.05 X	97.3	80.8
174	BAC	0.02	0.02	0.03	0.06 X	0.05 X	95.7	84.2
total		1.73	1.22	1.61	3.27 X	2.98 X	94.4	86.7

Supplementary Table 2. Statistics of the assembled papaya genome

Number of contigs	47,483
Total length of contigs (Mb)	271
N50 of contigs (kb)	11
Number of scaffolds	17,764
Total length of scaffolds (Mb)	370
N50 of scaffolds (Mb)	1
Number of anchored contigs	20,636
Total length of anchored and oriented contigs (Mb)	117
Total length of anchored not oriented contigs (Mb)	50
Number of anchored scaffolds	291
Total length of anchored and oriented scaffolds (Mb)	161
Total length of anchored not oriented scaffolds (Mb)	74

Supplementary Table 4. EST support for gene models in *Carica* genome. Counts represent the number of gene models with corresponding ESTs or Tribes with at least one corresponding EST.

	Carica genome	with EST	percent
Number of gene models	25312*	9648	38.1%
Number of Tribes	12958	5858	45.2%
Tribes *Carica* shares with *Arabidopsis*	6726	4871	72.4%
Tribes unique to *Carica*	5669	741	13.1%
Singleton Tribes unique to *Carica*	5314	681	12.8%

* This is one of the four gene models used for annotating the papaya genome.

Plant Communications
Research article

CellPress
Partner Journal

Pistachio genomes provide insights into nut tree domestication and ZW sex chromosome evolution

Salih Kafkas[1,15,*], Xiaokai Ma[2,3,15], Xingtan Zhang[2], Hayat Topçu[1], Rafael Navajas-Pérez[4], Ching Man Wai[5], Haibao Tang[2], Xuming Xu[2,6], Mortaza Khodaeiaminjan[1], Murat Güney[1], Aibibula Paizila[1], Harun Karcı[1], Xiaodan Zhang[5], Jing Lin[2], Han Lin[2], Roberto de la Herrán[4], Carmelo Ruiz Rejón[4], Jerson Alexander García-Zea[4], Francisca Robles[4], Coral del Val Muñoz[7,8], Agnes Hotz-Wagenblatt[9], Xiangjia Jack Min[10], Hakan Özkan[11], Elmira Ziya Motalebipour[1], Hatice Gozel[12], Nergiz Çoban[12], Nesibe Ebru Kafkas[1], Andrej Kilian[13], HuaXing Huang[2], Xuanrui Lv[2], Kunpeng Liu[2], Qilin Hu[2], Ewelina Jacygrad[14], William Palmer[14], Richard Michelmore[14] and Ray Ming[5,*]

	Siirt (female)	Bagyolu (male)
Assembly		
Number of scaffolds	50	28
Longest scaffold	57 779 128	62 820 281
Scaffold N50 (Mb)	38.7	39.8
Assembly length (Mb)	596.0	623.4
% of sequences anchored onto pseudochromosomes	99.3	99.9

TABLE 1

Overview of sequencing methods and characteristics

	Generation	Input	Template enrichment	Nature of sequencing template	Sequencing chemistry	Detection	Particularity	Reference
Sanger sequencing	1st	DNA	PCR	Pool	Primer extension, dideoxy termination	Radioactivity, fluorescence	Used for Human Genome Project	Sanger [1]
Maxam-Gilbert	1st	DNA	Sufficient template required	Pool	Chemical base-specific cleavage	Radioactivity		Maxam Gilbert [6]
454-Roche	2nd	DNA	Emulsion PCR	Amplified Clone	Primer extension one base at a time	Real-time chemiluminescence	Pyrosequencing	Margulies [11]
Polonator	2nd	DNA	Emulsion PCR	Amplified Clone	Oligonucleotide ligation	Fluorescence imaging		Shendure [17]
Solexa-Illumina	2nd	DNA	Bridge PCR	Amplified Clone	Single-base primer extension	Fluorescence imaging		Bentley [15]
Helicos	2nd	DNA	None	Individual DNA molecule	Single-base primer extension	Fluorescence imaging	Single molecule	Harris [19]
Applied Biotechnologies-Life Technologies	2nd	DNA	Emulsion PCR	Amplified Clone	Oligonucleotide ligation	Fluorescence imaging	Two-base encoding	McKernan [39]
Complete Genomics	2nd	DNA	Rolling circle replication	Amplified Clone	Oligonucleotide ligation	Fluorescence imaging	Proprietary system	Drmanac [20]
IonTorrent	2nd	DNA	Emulsion PCR	Amplified Clone	Primer extension one base at a time	pH measurement in semiconductor device	Non-optical	Rothberg [18]
Pacific Biosciences	2.5th	DNA	None	Individual DNA molecule	Primer extension	Real-time fluorescence imaging	Zero-mode wave guides	Eid [9]
Oxford Nanopore (Exonuclease)	3rd	DNA	None	Individual DNA molecule	Exonuclease digestion	Nanopore	Exonuclease nanopore	http://www.nanoporetech.com
Oxford Nanopore (Strand)	3rd	DNA	None	Individual DNA molecule	Direct measurement	Nanopore		http://www.nanoporetech.com