New Developments in 454 Sequencing and the Future of Sequencing at Roche

Mark Driscoll, Ph.D.
454 Life Sciences

454 SEQUENCING

IMPORTANT NOTICE

Intended Use

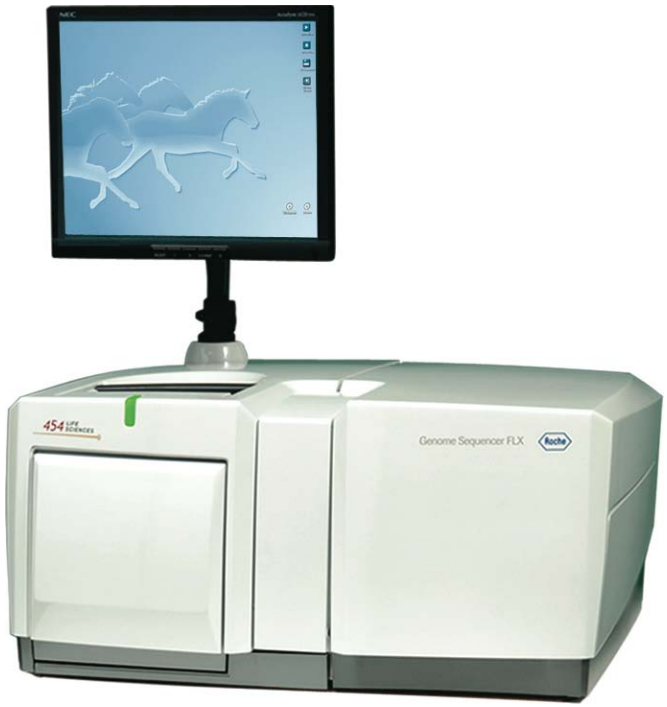
Unless explicitly stated otherwise, all Roche Applied Science and 454 Life Sciences products and services referenced in this presentation / document are intended for the following use:

For Life Science Research Only.

Not for Use in Diagnostic Procedures.

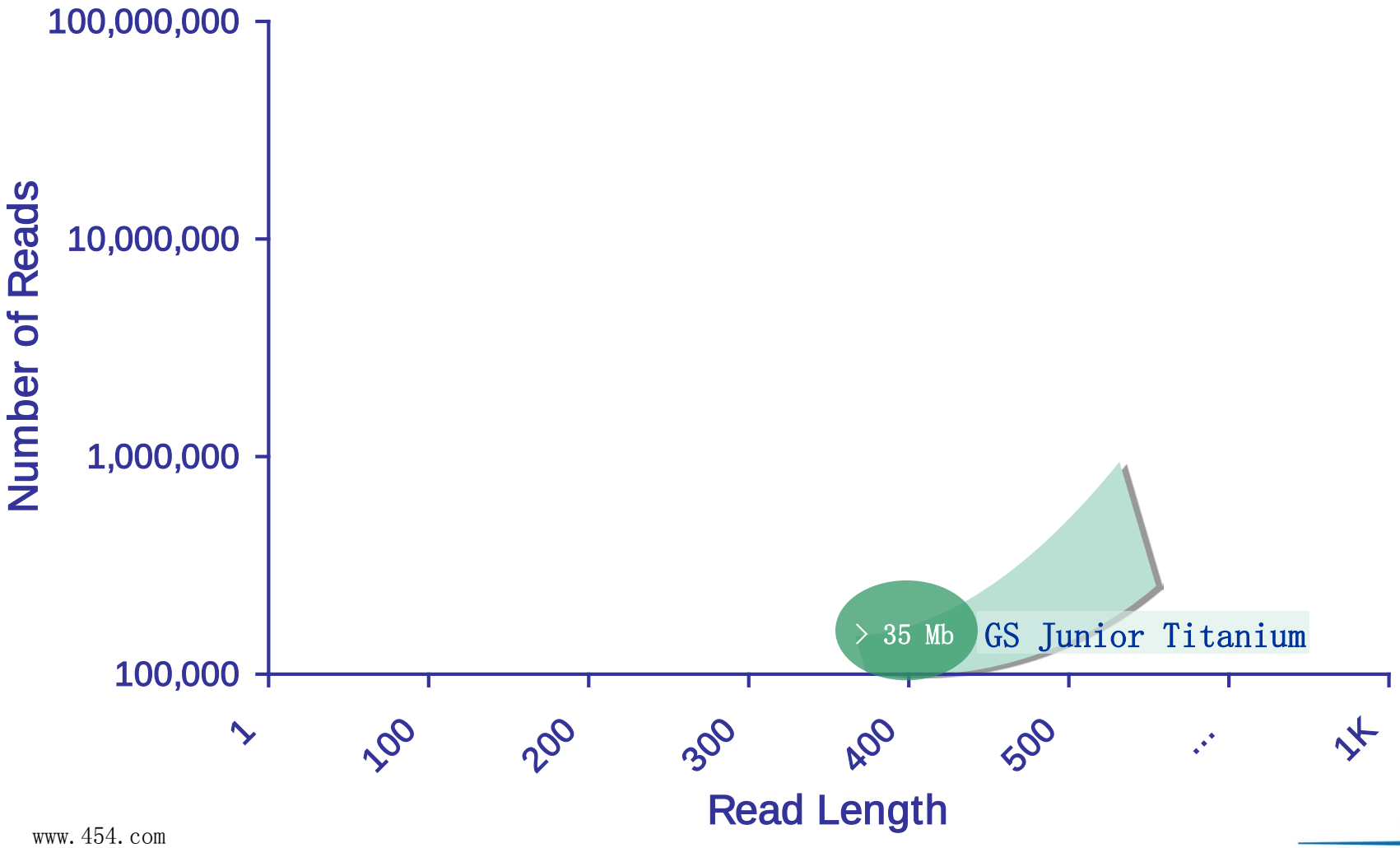
GS Junior

GS FLX+



Roche commitment to sequencing

Increasing throughput, read length, expanding portfolio

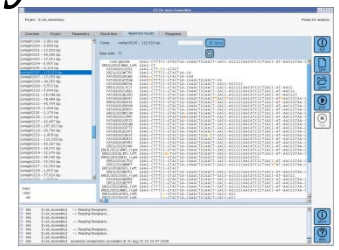


GS Junior System

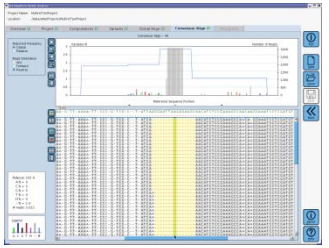
Next-gen sequencing & analysis- no big deal?

GS Junior System

An integrated solution from to sample to analysis



GS *De Novo* Assembler



Amplicon Variant Analyser

Day 1
Sample Preparation

Rapid Library Prep & emPCR set-up

Minimal additional lab equipment required

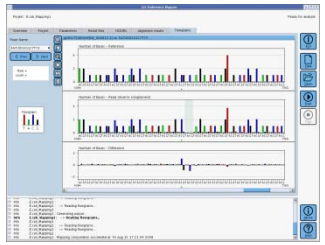
Day 2
Sequencing

emPCR enrichment

Thaw-and-go sequencing reagent cassette

Overnight sequencing- 10 hour run time

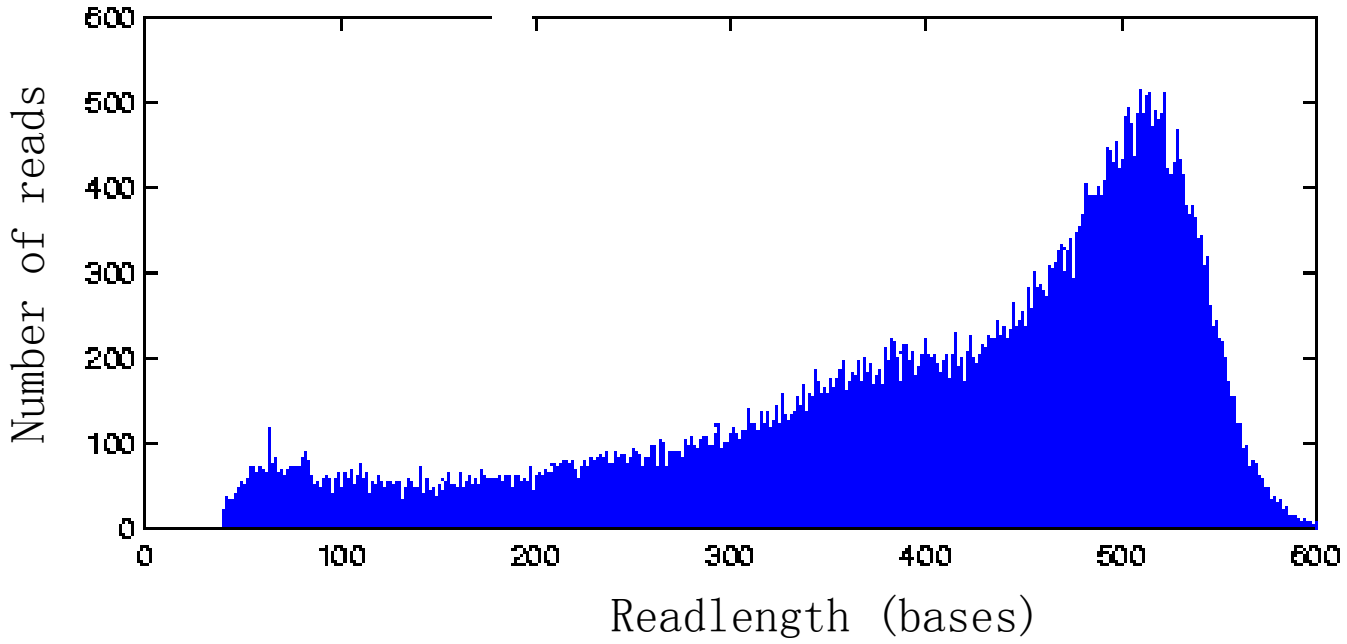
Day 3
Data Analysis



GS Reference Mapper

GS Junior Titanium System Read Length

- One GS Junior run produces reads from 50-600 or more in length
- Average is in 330-400 base range
- Read peak in the 450-550 base range



Performance Summary
GS Junior System

	GS Junior System
Throughput	35 million bases shotgun, 24 million bases amplicon (approx.)
Avg. Read Length	400 bases (approx.)
HQ Reads per Run	100,000 shotgun, 70,000 amplicon (approx.)
Accuracy	Q20 read length at 400 bases (99% accuracy at 400 bases)
Run Time	10 hours sequencing, 2 hours data processing
Sample Input	Purified gDNA, amplicons, cDNA, depending on application

*Per run specifications is for shotgun libraries, and can vary based on the organism and genomic content. Reference organism is E. coli.

Demonstrated Success with 454 Sequencing Systems

Enabling breakthrough genomic discoveries

1200+

peer-reviewed publications

de novo Transcriptome Epigenetics

Metagenomics SeqCap

Resequencing Rare Variants Det.

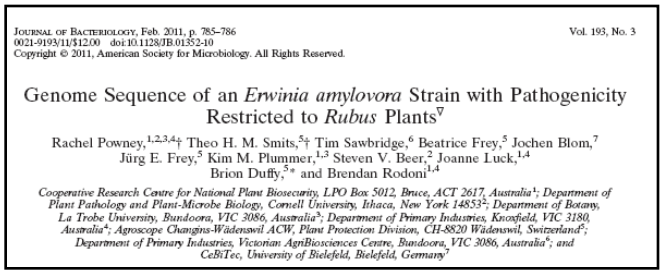
Chip-Seq **Genotyping** Small RNA

Ancient DNA Methylation Expression Tags

de novo **Whole Genome**

www.454.com

Recent GS Junior System Publications

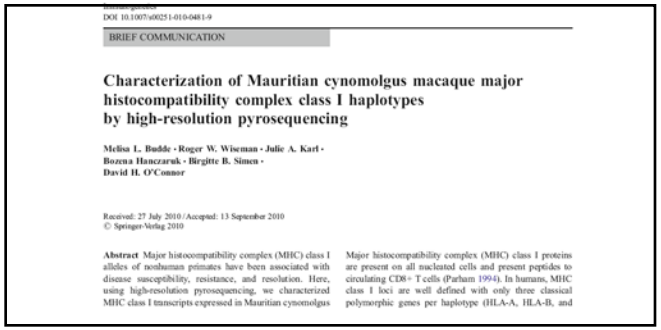


JOURNAL OF BACTERIOLOGY, Feb. 2011, p. 785–786 Vol. 193, No. 3
0021-9193/11/$12.00 doi:10.1128/JB.01352-10

Genome Sequence of an *Erwinia amylovora* Strain with Pathogenicity Restricted to *Rubus* Plants[▽]

Rachel Powney,[1,2,3,4]† Theo H. M. Smits,[5]† Tim Sawbridge,[6] Beatrice Frey,[5] Jochen Blom,[7] Jürg E. Frey,[5] Kim M. Plummer,[1,3] Steven V. Beer,[2] Joanne Luck,[1,4] Brion Duffy,[5]* and Brendan Rodoni[1,4]

Cooperative Research Centre for National Plant Biosecurity, LPO Box 5012, Bruce, ACT 2617, Australia[1]; Department of Plant Pathology and Plant-Microbe Biology, Cornell University, Ithaca, New York 14853[2]; Department of Botany, La Trobe University, Bundoora, VIC 3086, Australia[3]; Department of Primary Industries, Knoxfield, VIC 3180, Australia[4]; Agroscope Changins-Wädenswil ACW, Plant Protection Division, CH-8820 Wädenswil, Switzerland[5]; Department of Primary Industries, Victorian AgriBiosciences Centre, Bundoora, VIC 3086, Australia[6]; and CeBiTec, University of Bielefeld, Bielefeld, Germany[7]



DOI 10.1007/s00251-010-0481-9

BRIEF COMMUNICATION

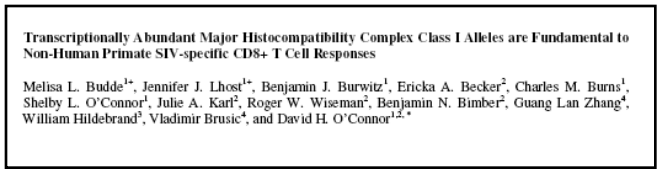
Characterization of Mauritian cynomolgus macaque major histocompatibility complex class I haplotypes by high-resolution pyrosequencing

Melisa L. Budde · Roger W. Wiseman · Julie A. Karl · Bozena Hanczaruk · Birgitte B. Simen · David H. O'Connor

Received: 27 July 2010 / Accepted: 13 September 2010

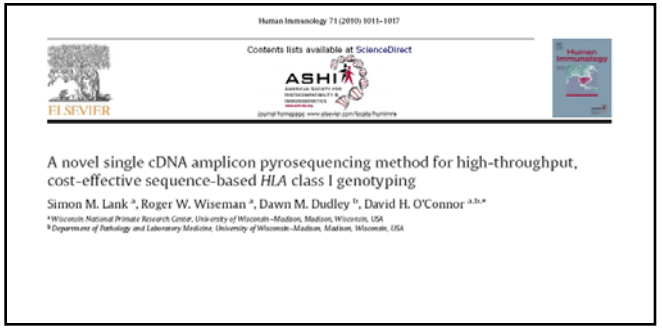
Abstract Major histocompatibility complex (MHC) class I alleles of nonhuman primates have been associated with disease susceptibility, resistance, and resolution. Here, using high-resolution pyrosequencing, we characterized MHC class I transcripts expressed in Mauritian cynomolgus

Major histocompatibility complex (MHC) class I proteins are present on all nucleated cells and present peptides to circulating CD8+ T cells (Parham 1994). In humans, MHC class I loci are well defined with only three classical polymorphic genes per haplotype (HLA-A, HLA-B, and



Transcriptionally Abundant Major Histocompatibility Complex Class I Alleles are Fundamental to Non-Human Primate SIV-specific CD8+ T Cell Responses

Melisa L. Budde[1*], Jennifer J. Lhost[1*], Benjamin J. Burwitz[1], Ericka A. Becker[2], Charles M. Burns[1], Shelby L. O'Connor[1], Julie A. Karl[2], Roger W. Wiseman[2], Benjamin N. Bimber[2], Guang Lan Zhang[4], William Hildebrand[3], Vladimir Brusic[4], and David H. O'Connor[1,2,*]



Human Immunology 71 (2010) 1011–1017

Contents lists available at ScienceDirect

ELSEVIER ASHI

A novel single cDNA amplicon pyrosequencing method for high-throughput, cost-effective sequence-based *HLA* class I genotyping

Simon M. Lank [a], Roger W. Wiseman [a], Dawn M. Dudley [b], David H. O'Connor [a,b,*]

[a] Wisconsin National Primate Research Center, University of Wisconsin–Madison, Madison, Wisconsin, USA
[b] Department of Pathology and Laboratory Medicine, University of Wisconsin–Madison, Madison, Wisconsin, USA

Whole Genome *de novo* Sequencing Of Microbes

Targeted Amplicon Sequencing

Virology/Immunology

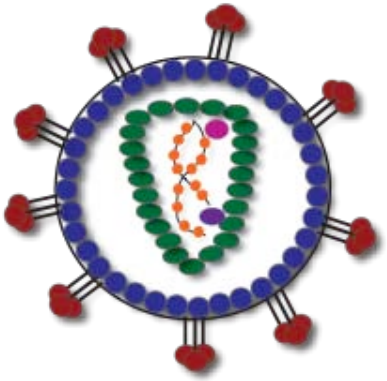
Gene Expression

SIV Evolution During Immune Response

Sequencing amplicons using the GS Junior System

Goals: 1. Follow changes in GAG gene as virus evolves to evade immune resp

2. Find genome-wide mutations in viral pool



Simian Immunodeficiency Virus

Rhesus macaque

Acknowledgements

- Ben Burwitz in Dave O'Connor's lab, Univ. of Wisconsin, supplied the data

O'Connor Lab

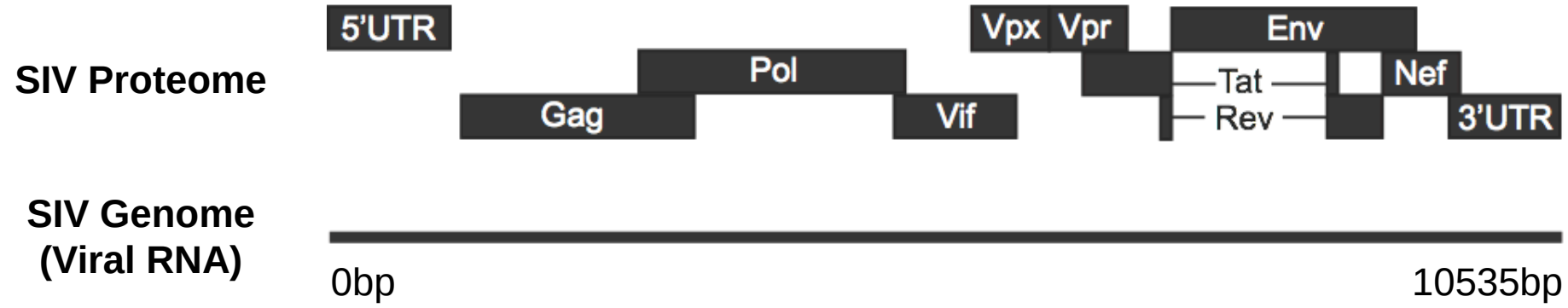
Ben Burwitz
Roger Wiseman
Shelby O'Connor
Dawn Dudley
Julie Karl
Simon Lank
Charlie Burns
Ericka Becker
Ben Bimber
Dave O'Connor

Watkins Lab

Jonah Sacha
Matt Reynolds
Nick Maness
Nancy Wilson
David Watkins

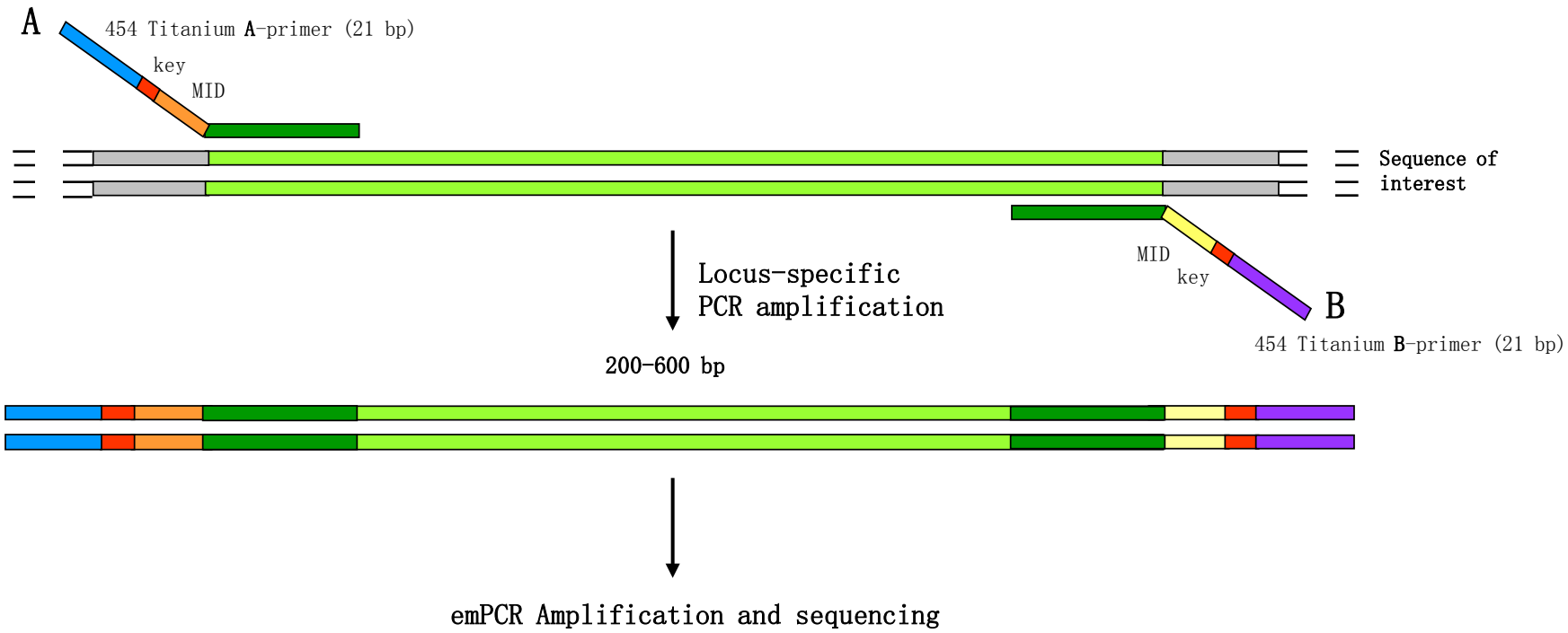
SIV Genome Sequencing

Two different amplicon approaches using the GS Junior System



Amplicon Sequencing- Basic Amplicon Design

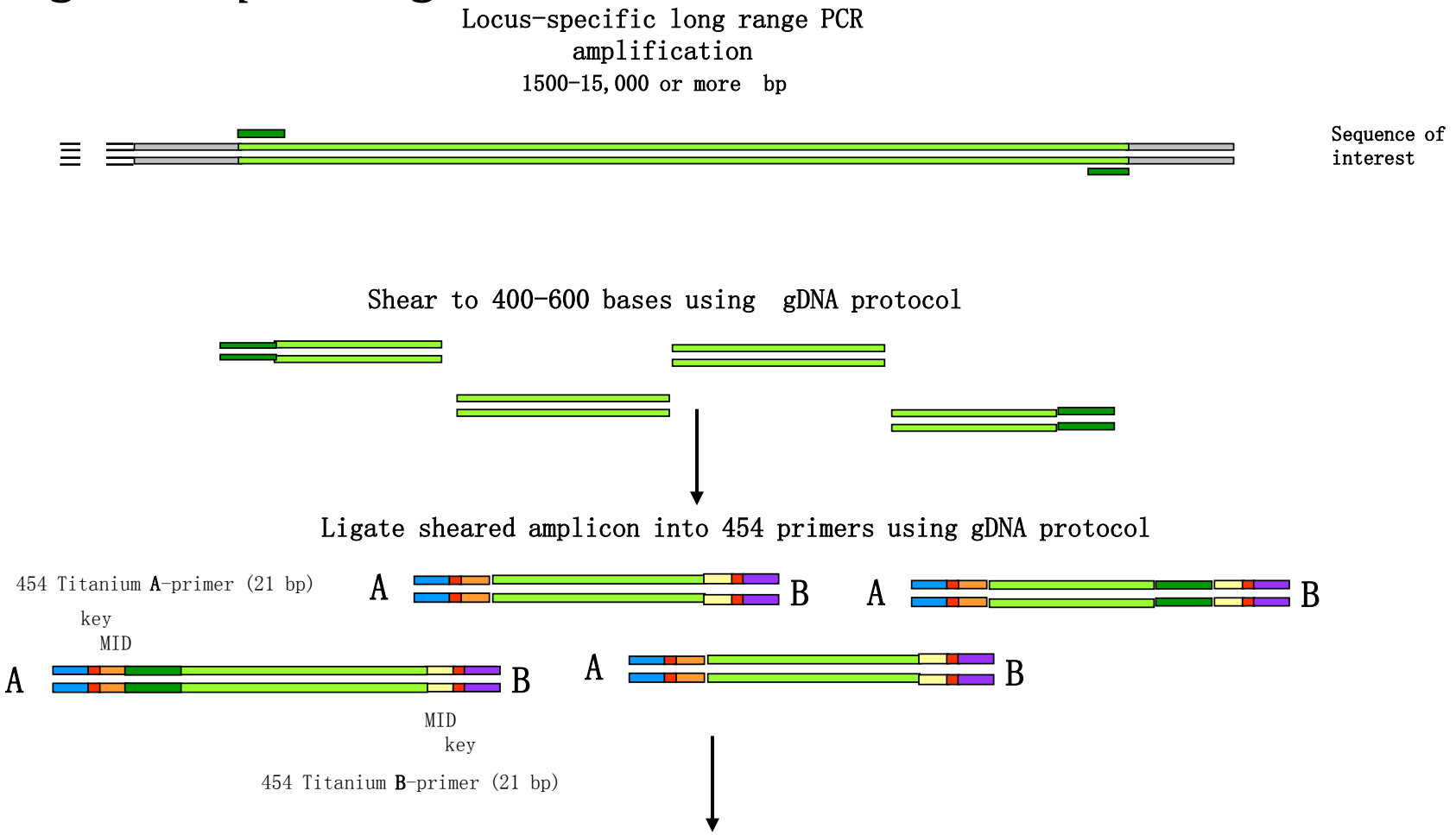
454 amplicon design using tailed primers



- Long reads required to sequence through the locus specific primer, enable haplotyping over longer distances
- 100s to 1000s of amplicon clones sequenced simultaneously

Amplicon Sequencing- Long Range Amplicons

using long range amplicons for whole viral or other genomic region sequencing



SIV Genome Sequencing – Direct Amplicon

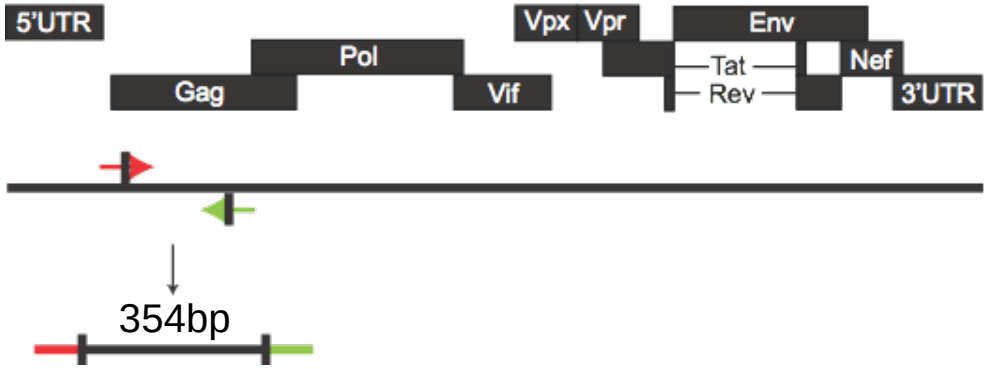
Roche

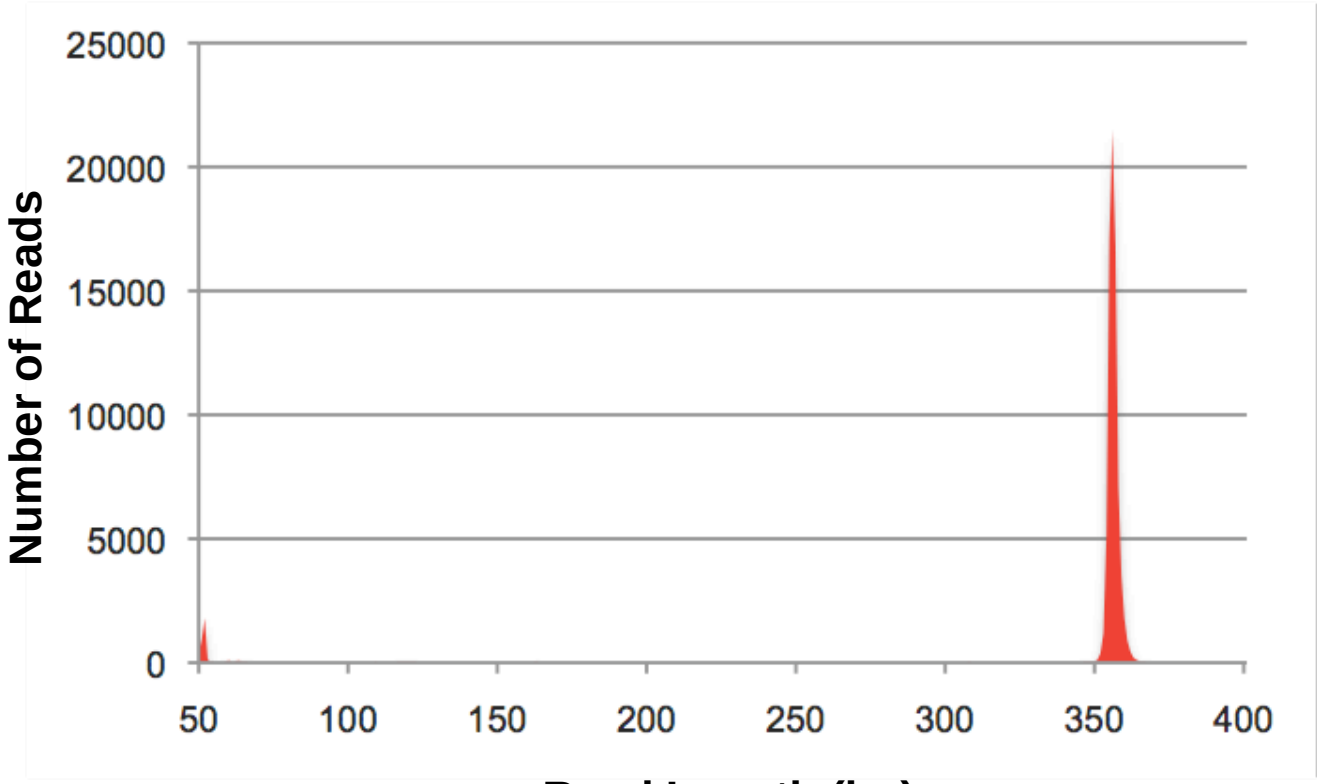
Single Run Results

28 different Samples

82,079 Reads

356 base Median Read Length

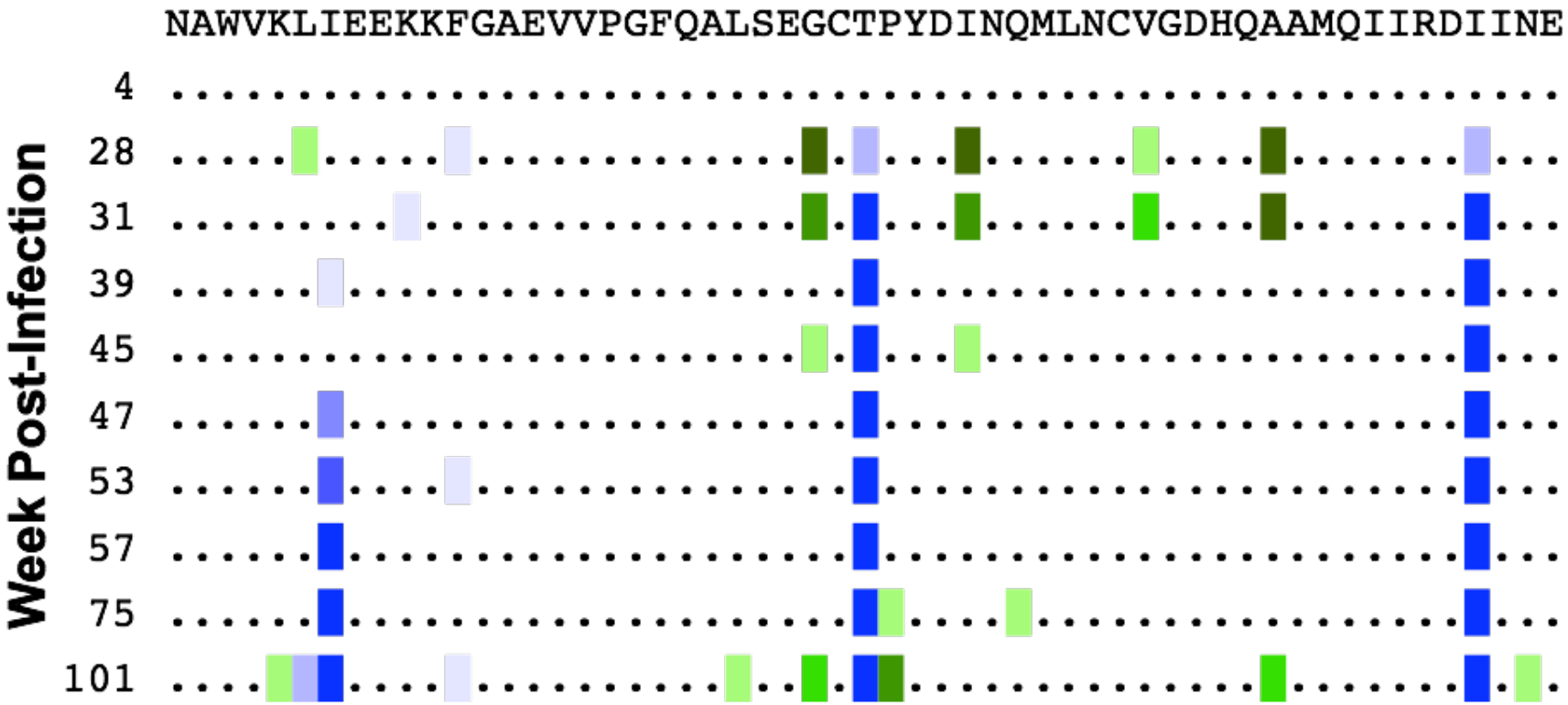




454 SEQUENCING

SIV Gag evolution

Protein coding changes in response to immune system pressure

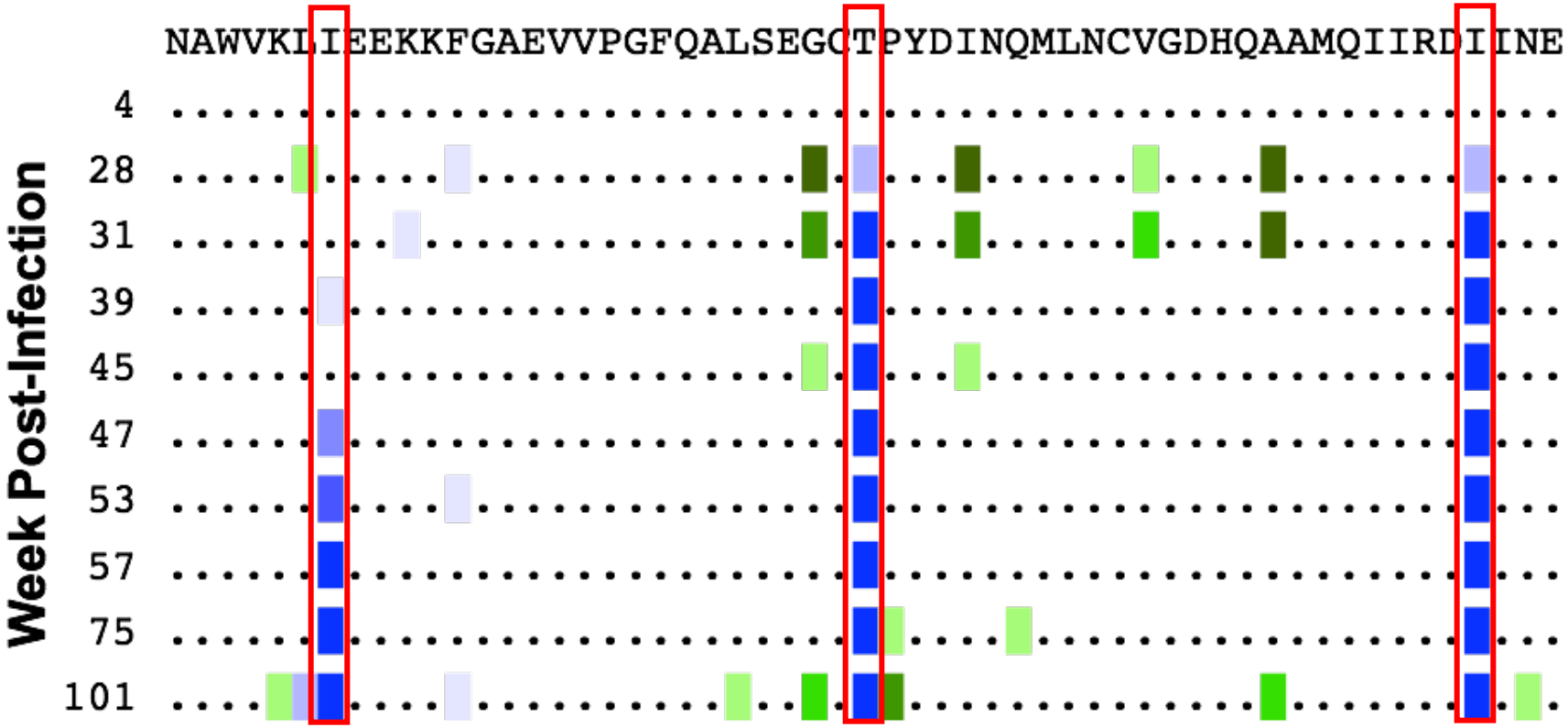


Feature	Percent
Nonsynonymous Mutation	1-3
	3-5
	5-10
	10-30
	30-100
Synonymous Mutation	1-3
	3-5
	5-10
	10-30
	30-100

Mutations in the SIV protein Gag affect viral fitness- Gag protein is the 'particle making machine'

SIV Gag evolution

Protein coding changes in response to immune system pressure

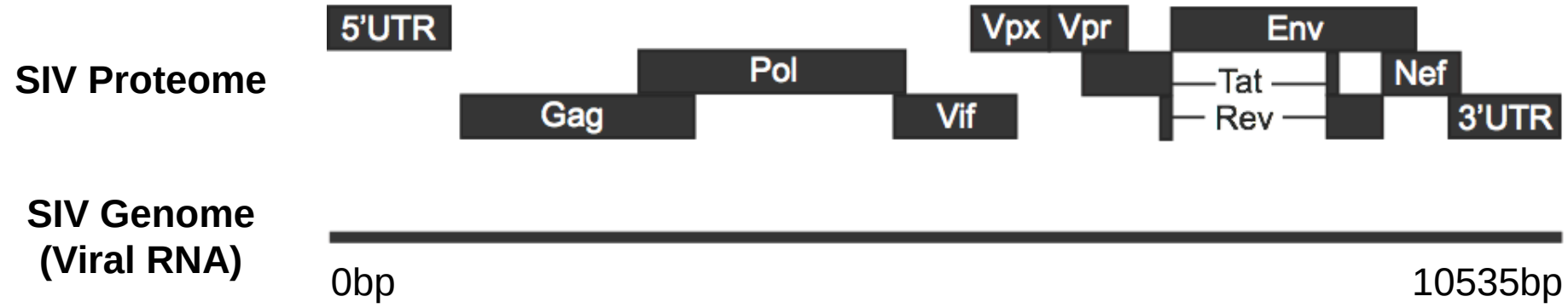


Feature	Percent
Nonsynonymous Mutation	1-3
	3-5
	5-10
	10-30
	30-100
Synonymous Mutation	1-3
	3-5
	5-10
	10-30
	30-100

Mutations in the SIV protein Gag affect viral fitness- Gag protein is the 'particle making machine'

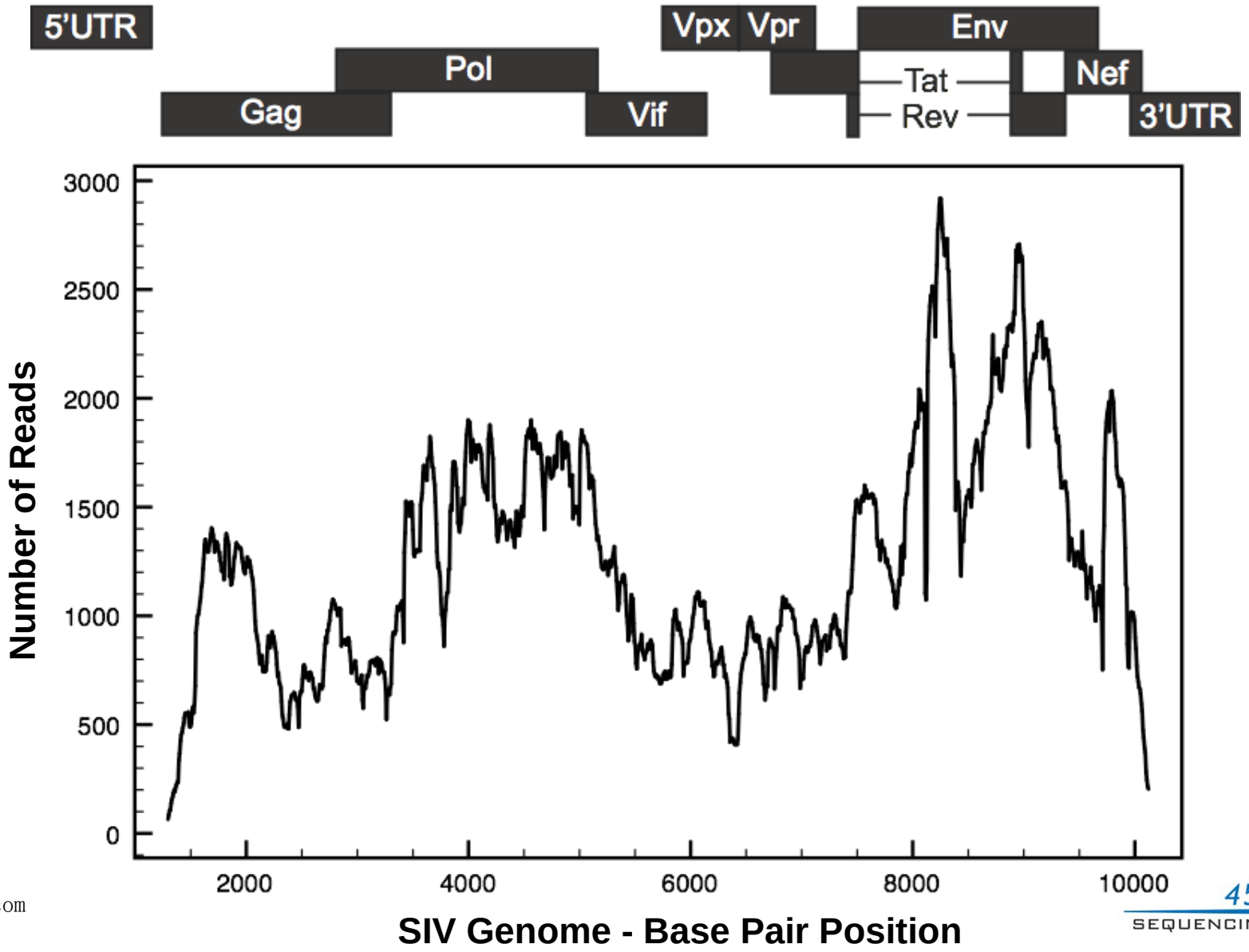
SIV Genome Sequencing

Two different amplicon approaches using the GS Junior System



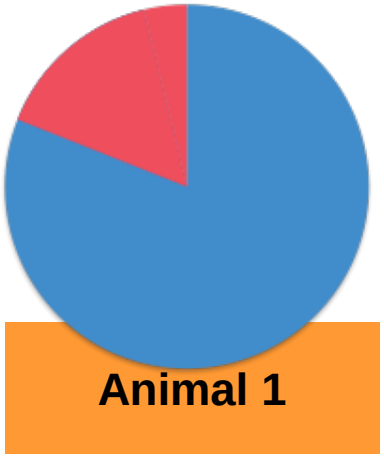
SIV Whole Genome Coverage

Tiled amplicons covering the entire genome



GS Junior System vs. Sanger Sequencing

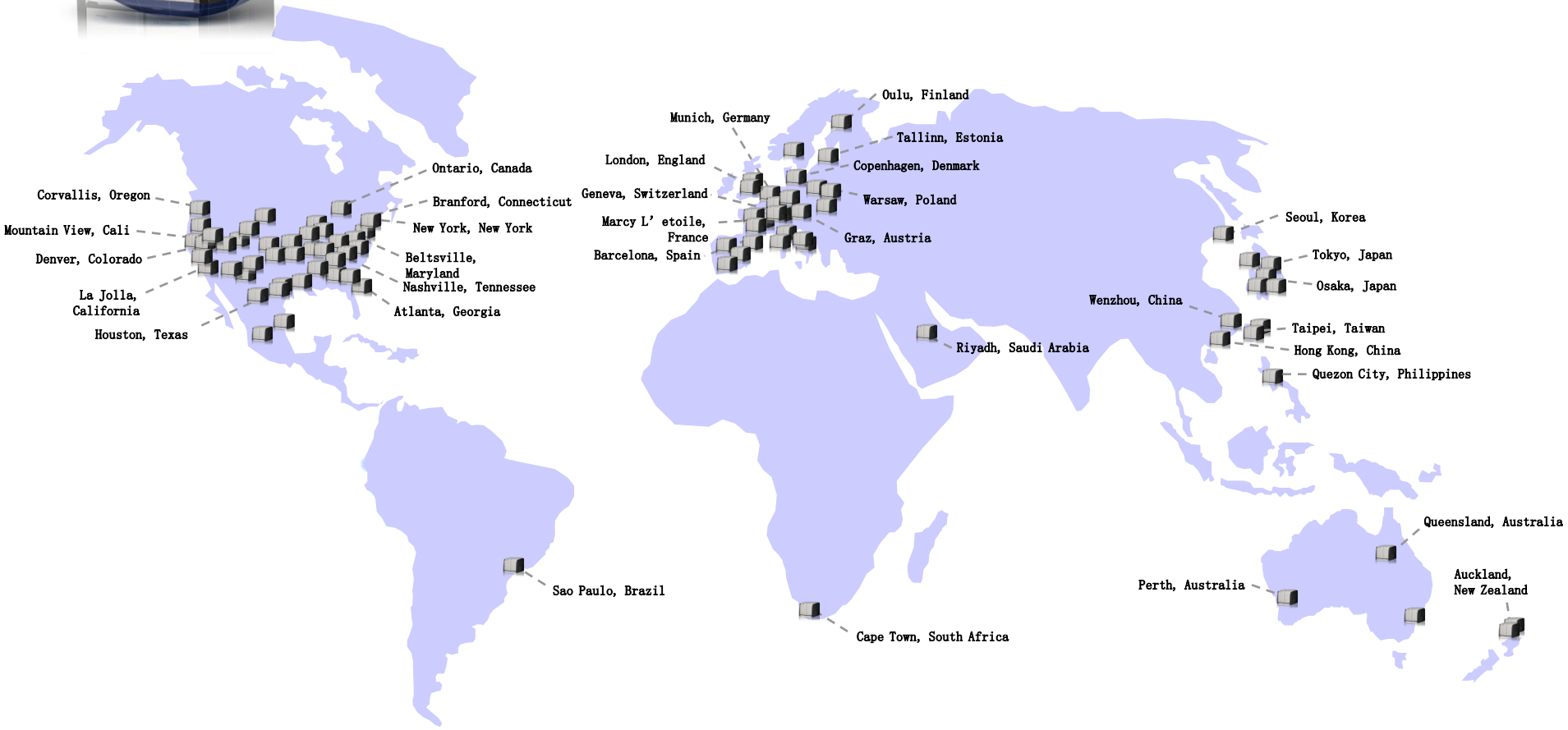
More long reads= more mutations found!



Category	Average
# of mutations detected by Roche/454 at 5% or greater	144
Mutation detected only by Roche/454	76.5%
Mutation detected by both Roche/454 and Sanger	33.5%

Where In the World is GS Junior?

In hundreds of labs worldwide!



www.454.com

Note: Placements shown are a representative sample.

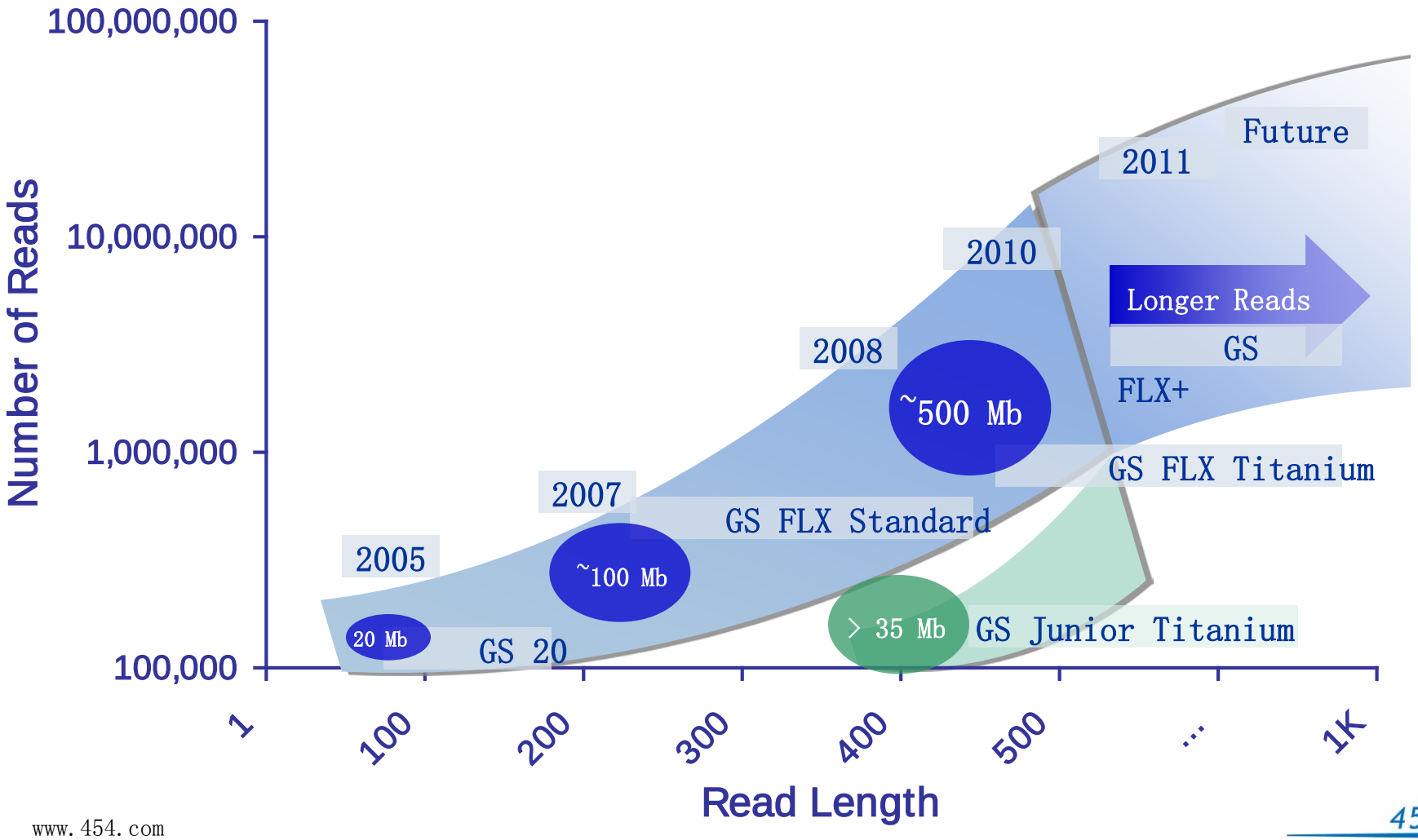
Research Assays Menu

For both GS Junior and GS FLX Systems

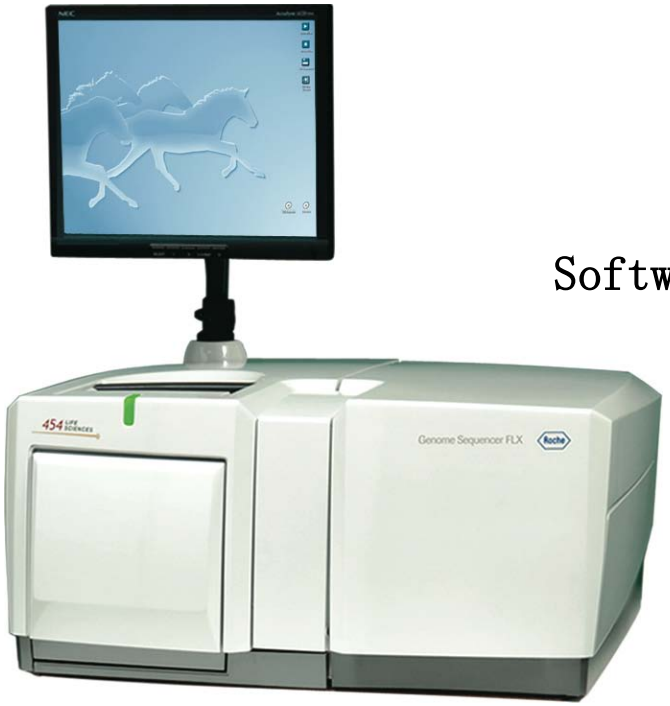
Gene Target	Application	Availability
HLA class 1 & 2 GS GType HLA Primers	• High & medium resolution genotyping kits • Association with wide variety of autoimmune diseases, cancers, infectious pathogens	Q2 2011
TET2 (CBL, KRAS)	• Primary association with leukemia • Developed in collaboration with the Munich Leukemia Laboratory	In development
RUNX1	• Same as above	In development
EGFR	• Association with lung cancer and other cancers	In development
16S	• Metagenomics studies - bacterial identification	Planned
VDJ	• Immune repertoire monitoring	Planned
BRCA	• Association with breast cancer, ovarian cancer	Planned
Pathogen Detection	• Assay for suppression of host sequence • Developed by Lipkin Lab at Columbia University	Planned

Roche commitment to sequencing

Increasing throughput, read length, expanding portfolio



GS FLX+ System



What’ s New:

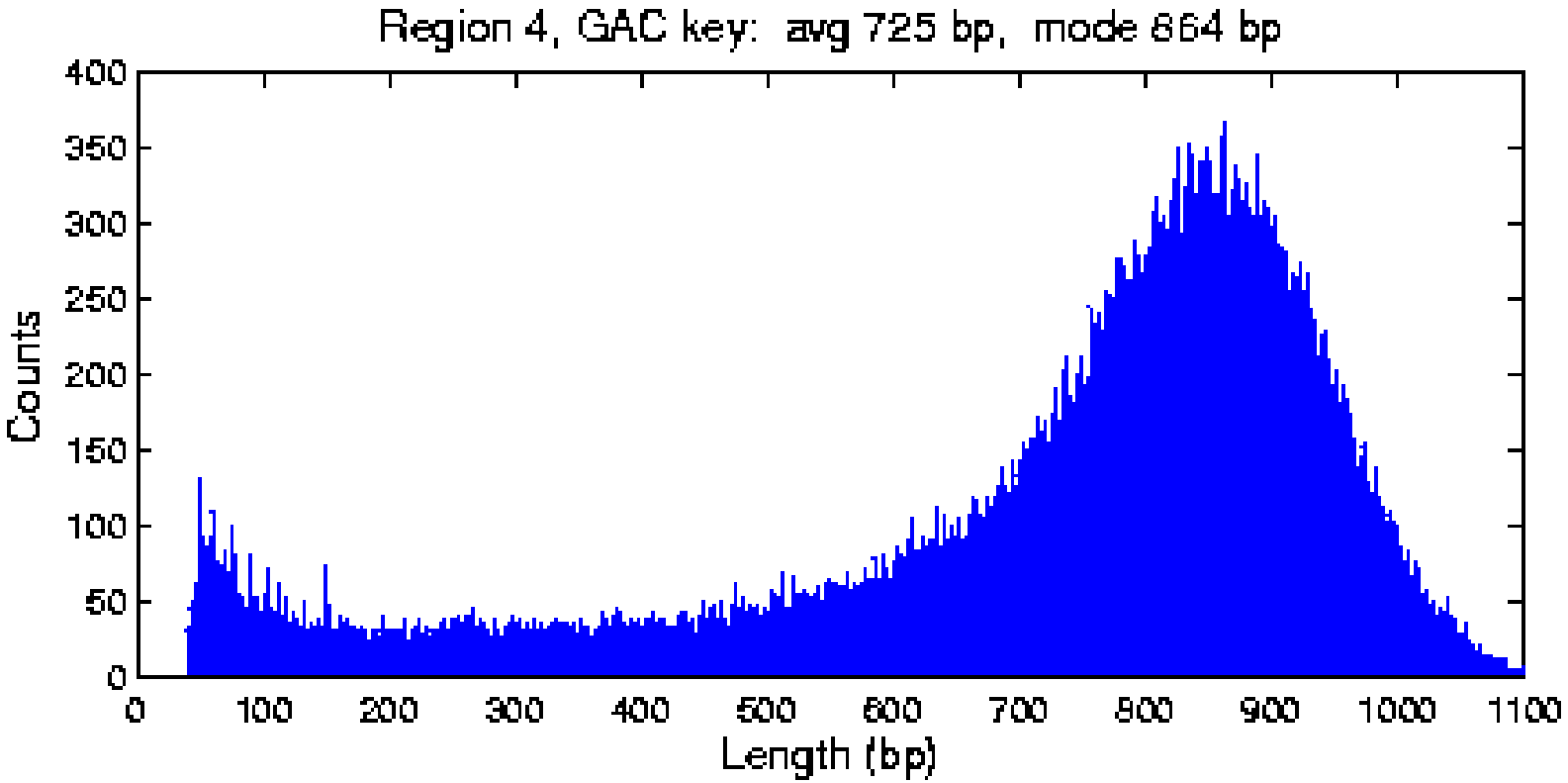
Hardware Upgrade (new fluidics)

Software Upgrade (added capability for long reads)

New Kits- long read capability

Backward compatible with GS FLX Titanium chemistry!

GS FLX+ Sequencing Run Example



We can show

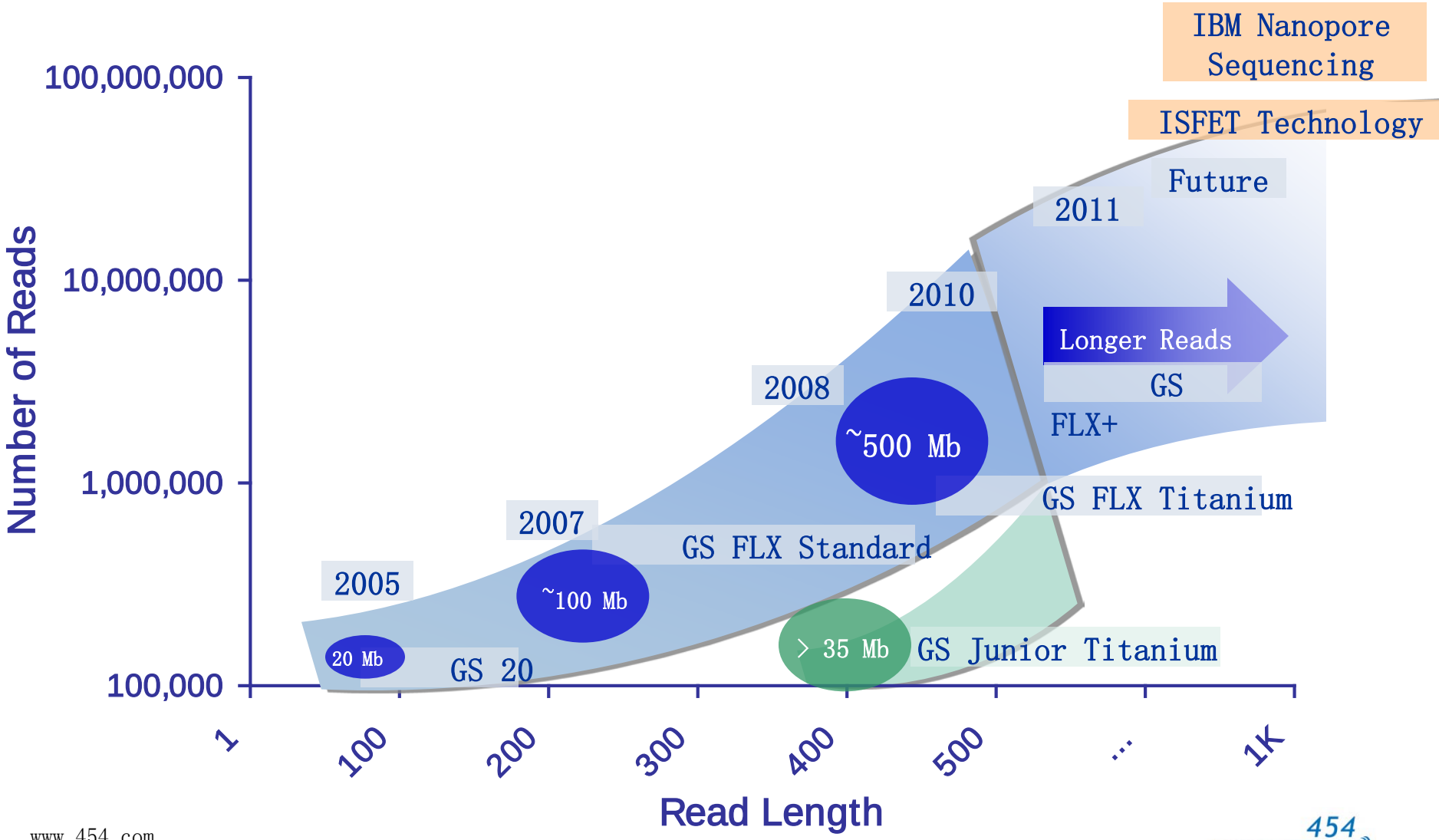
Budgie Bird *De Novo* Assembly
How do GS FLX + reads help assembly?

Assembly Metrics	GS FLX Titanium (w/ Paired End)	GS FLX+ (w/ Paired End)	Improvement
Sequence Depth	6X	6X	-
Number of Bases	8,018,686,780	8,019,891,335	Equal number of bases
Scaffold Metrics			
Avg. Scaffold Size	22.5 Kb	23.1 kb	3%
N50 Scaffold Size	1.9 Mb	2.5 mb	27%
Largest Scaffold Size	14.0 Mb	15.6 mb	11%
Contig Metrics			
Number Of Contigs	418 k	302 k	-28% (fewer is better)
Avg. Contig Size	2.3 kb	3.3 kb	42 %
N50 Contig Size	3.2 kb	5.2 kb	60 %
Largest Contig Size	39 kb	57 kb	47 %

The Future of 454 Sequencing

Roche commitment to sequencing

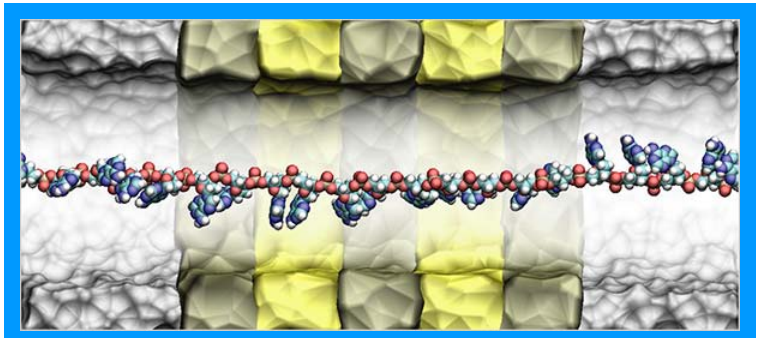
Increasing throughput, read length, expanding portfolio



Nanopore Sequencing
IBM DNA Transistor Technology

- Single molecule sequencing; no amplification needed
- Simple sample preparation
- Low reagent/disposable cost
- Very, very long read lengths
- Extremely fast (~1000 bp/sec/nanopore) and scalable up to throughput in the Tb range
- Use electrical base detection (no optics and therefore inexpensive)

→ Target- $100 genome



IBM DNA Transistor

DNA molecules are read as they rachet through nanopore-sized holes in a silicon chip.

454 Development Program
Pushing the limits of sequencing

GS Junior and GS FLX

- **Performance enhancements-** Improved data quality, read length and throughput
- **Dedicated assays-** targeting useful genomic panels
- **Automation-**limit hands-on time requirement
- **Software-**Newbler, AVA, GS Mapper
- **Sanger-like read lengths** - simplified, powerful analysis

GS FLX

- Extend read length to match Sanger
- Substantially improve complex genome assembly

Nanopore Sequencing

- Genome in minutes
- New technology
- $100 genome

For life science research only. Not for use in diagnostic procedures.