#### How to Multiple Align Huge Number of Pyrosequencing Reads

0

#### Fahad Saeed and Ashfaq Khokhar

Department of Electrical and Computer Engineering, University of Illinois at Chicago

#### Osvaldo Zagordi, and Niko Beerenwinkel

Department of Biosystems Science and Engineering, ETH Zurich



#### Overview

- Genome Sequencing with Pyrosequencing (454 GS20 Platform)
- Need for Multiple Alignments
- Problem Statement and Assumptions
- Algorithmic approach
- Quality Results
- Ongoing works
- Conclusions and Future work

#### A look at the 454 sequencing technology



## Genome Sequencing with pyrosequencing

- Fast and Cheap Sequencing: an important goal
- Pyrosequencing: attractive alternative to current sequencing techniques

	Sanger	454/Roche
bps per run	~10 <sup>5</sup>	~10 <sup>8</sup>
read length	700-1000	~400
cost per run	~1000 \$	~15000 \$
cost per Mbp	10K \$	100 \$
accuracy	high	low (in-dels)

#### Key Issues with pyrosequencing

- Read Length: on average 250-400bp
- Orientation: Original and complement (Ignored for this work)
- Errors
  - Insertions: ~36%
  - Deletions: ~27%
  - Ambiguous Bases: ~21%
  - Substitutions: ~16%

#### **Need for Alignments**

- Required as a preprocessing step for number of procedures.
- For example, from reads to haplotypes:



#### Pairwise alignment

- If reference genome is available, then pairwise alignments can be done
- But don't work :
  - Primarily because of insertions due to pyrosequencing

read\_0159|beg|1|/ength|250/1-250 read\_0210|beg|2|length|254/1-254 read\_0123|beg|2|length|250/1-250 read\_1168|beg|2|length|256/1-256 read\_1640|beg|2|length|253/1-253 read\_0543|beg|4|length|253/1-253 read\_0583|beg|5|length|256/1-256 read\_0655|beg|6|/ength|250/1-250 read\_0848|beg|9|/ength|253/1-253 read\_1866|beg|7|length|249/1-249 read 0307|beg|7|length|249/1-249 read\_1253|beg|8|/ength|250/1-250 read\_0735|beg|8|/ength|253/1-253 read\_0181|beg|9|/ength|254/1-254 read 0282|beg|11|length|248/1-248 read\_0746|beg|11|length|250/1-250 read 1895|beg|12|length|256/1-25 read\_1608|beg|14|length|252/1-25. read\_1124|beg|15|length|253/1-25 read\_0223|beg|15|length|254/1-254 read\_0383|beg|15|length|250/1-250 read\_1997|beg|15|length|257/1-25 read\_1176|beg|16|length|256/1-256 read\_1170|beg|16|length|257/1-25 read\_1543|beg|17|length|256/1-256 read\_1534|beg|20|length|252/1-25. read\_0292|beg|19|length|255/1-25 read\_0689|beg|19|/ength|252/1-25. read\_0303|beg|20|length|254/1-254 read\_1450|beg|21|length|249/1-24 read\_0754|beg|23|length|251/1-25 read\_0068|beg|23|length|251/1-25 read\_0390|beg|23|length|253/1-25 read\_1594|beg|24|length|254/1-254 read\_1492|beg|25|length|260/1-260 read\_1489|beg|25|length|252/1-25.

<>	80	90	100	110	120	130	140	150	160
	AGATACAGGAGCA	GATGATACA	GTATTAG AGA	AATGAATTT	GCCAGGAAGA	TGGAAACC AA	AATGATAGGO	GGGAATTGGG	AGGTTT
	ATTAGATACAGGA	GC-GATGAT	ACAGTACTTAG	- AGAAATGA	ATTTGCCAGG.	AA <mark>gatgg</mark> aaaci	C - AAAATGAT	AGGGGGGAATT	GGAGGT
	T T AG A T A C A G G A G	CAGATGATA	CAGTATTAG - A	GAAATGAAT	TTGCCAGGAA	GATGGAAACC-	AAAATGATAG	GGGGAATTGG	AGGTTT
	TTAGATACAGGAG	CAGATGATA	CAGTATTAG - A	G - AATGAAT	TTTGCCAGGA.	A <mark>gatgg</mark> aaacc	AAAATGATA	GGGGGGAATTG	GAGGTT
	TAGATACAGGAGC	AGATGATAC	AGTATTAG - AG	ΑΑΑ <mark>ΤΘ</mark> ΑΑΤΤ	TGCCAGGAAG.	A <mark>tgg</mark> aaacc/	AAA <mark>TG</mark> ATAGG	GGGAATTTGG	AGGTTT
	TAGATACAGGAGC	AGATGATAC	AGTATTAG · A ·	AAATGAATT	TGCCAGGAAG.	AT <mark>gg</mark> aaacc-a/	AAATGATAGG	IGGG <mark>AATTGGT</mark>	AGGTTT
	AGATACAGGAGCA	GATGATACA	GTATTAG - AGA	AATGAATTT	GCCACGGAAG.	A <mark>tgg</mark> aaacc <mark>g</mark> a/	AAATGATAGG	GGGAATTTGG	- GGTTT
	TAGATACAGGAGC	AGATGATAC	AGTATTAG - AG	AAATGAATT	TGCCAGGAAG.	ATGGAAACC - AA	AAATGATAGG	GGGAATTGGA	GGTTTT
	AGATACAGGAGCA	GATGATACA	GTATTAG - AGA	AATGAATTT	GCCACGGAAC	GATTGGTAAACI	C-AAAATGAT	AGGGGGGAATT	GGAGGT
	TAGATACAGGAG	CAGATGATA	CAGTATTAG - A	GAAATGAAT	TTTGCC-GGA.	AGATGGAAACC	AAAATGATA	GGGGGGAATTG	GAGGTT
	AGATALAGGAGL	AGATGATAL	AGTATTAG AG	AAATGAATT	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	A-GGAAALL-A	AATGATAGG	GGGAATIGGA	GGITT
	TTAGATACAGGAGLA	CAGATGATACA	CAGTATTAG A	CAAAT CAATT	TTOCCAGOAAOA	ATGGAAACCAAA)	AAAATAGATAG	GGGGGAATTGGAG	AGGTTT
	AGATACAGGAGC.	GATGATACA	GTATTAG. AGA	AAATGAATT	TACCAGGAAC	SATGGAAACC.	AAATGATAG	GGGGGAATTGG	AGGTTT
8	AGAT - CAGGAGCA	GATGATACA	GTATTAG AGA	AATGAATTT	OCCAGGAACG	ATGG - AACC - AA	AATGATAGO	GGGAATTGGA	GGTTTT
0	AGATACAGGAGCA	GATGATACA	GTATTAG - AGA	AATGAATTT	GCCAGGAAGA	TGGAAACC - AAA	AATGATAGGG	GGAATTGGAG	GTTTTA
6	TAGATACAGGAGC	AGACTGATA	CAGTATTAG - A	GAAATGAAT	TTOCCAGGAA	GATGGAAACC -	AAAATGATAG	GOGGAATTOG	AGGTTT
2	AGATACAGGAGCA	CGATGATAC	AGTATTAG - AG	AAATGAATT	TOCCAGGAAG.	ATGGAAACC - AA	AAATGATAGG	GGGAATTGGA	GGTTTT
3	TTAGATACAGGAG	CAGATAGAT	ACAGTATTAG -	Α <mark>ς</mark> ΑΑΑ Τς ΑΑ	TTTGCCAGGA.	A <mark>gatgg</mark> aaacc	AAAATGATA	GGGGGGAATTG	GAGGTT
4	TAGATACAGGAGC	AGATGAT - C	AGTATTAG - AG	AAATGAATT	TGCCAGGAAG.	ATGGAAACC - AA	AAATGATAGG	GGGAATTGGA	GGTTTT
ο ,	A <mark>gatacaggagca</mark>	GATGATACA	GTATTAG - AGA	AATGAATTT	G C C A G G A A G A	T <mark>gg</mark> aaacc-aa/	AATAGATA-G	GGGAATTGGA	GGTTTT
7	TATTAGATACAGG	AGTCAGATG	ATACAGTATTA	G T A G A A A T G	AATTTGCCAG	GAAGATGGAAAI	CC - AAAATGA	I- AGGGGGGAAT	TGGAGG
6	AGATACAGGAGCA	GATGATACA	GTATTAG - AGA	AATGTAATT	TOCCAGGAAG.	A TGGAAACC · A	AAATGATAGG	GGGTAATTGG	AGGTTT
7	TAGATACAGGAGC	AGACTGATA	COGTATTAG - A	GAAATGAAT	TTAGCCAGGA.	A <mark>gatggaaacc</mark>	AAAATGATA	GGGGGGAATTG	GAGGTT
6	TTAGATACAGGAG	CAGATGATA	CACGTATTAG -	AGAAAATGA	ATTTGCCAGG.	AAGATGGAAACI	CCAAAATGAT	AGGGGGGAATT	GGAGGT
2	3 A T A C AG G AG C AG	ATGATACAG	TATTAG - AGAA	ATGAATTTG	CCAGGAAGAT	GGAAACC - AAA	ATGACTAGGG	GGAATTGGAG	GTTTTA
0	AGATALAGGAGTU	AGATGATAL	AGTATTAG - AG	AAATGAATT	TOCCAGGAAG.	ATGGAAACC - AA	AAATGATAGG	GGGGATTGGA	GGTTTT
4	AGATACAGGAGCA	ACATCATACA	AUTATTAG A A	AAATGAATT	CCAGGAAGA	TechAACCCAA	AATGATAGG	GGGAATTGGA	GTTTTA
9	ATACAGGA, CAG.	TGATACAGT	ATTAG. AGAAA	CTGAATTTG	CCAGGAAGAT	GAAACC. AAAA	AT. ATAGGGG	GAATTGGAGG	TTTTAT
1	SATACAGGAGCAG	ATGATACAG	TATTAG. AGAA	ATGAATTTG	CCAGGAAGAT	GGAAACC. AAAA	ATGATAGGGG	GGAATTGGAG	GTTTTA
1	ATACAGGAGCAG	ATGATACGA	GTATTAG - AGA	AATGAATTT	GCCAGGAAGA	TGGAAACC-AAA	AATGATAGGG	GGAATTGGAG	GTTTTA
3	AGATACAGGAGCA	GATGATACA	GTATTAG - AGA	AATGAATTT	GCCAGGAAGA	TGGAAACC - AA	AATGATAGGG	GGGAATTGGA	GG . TTT
4	AGATACAGGAGCA	GATGATACA	GTATTAG - AGA	AATGAATTT	GCCAGGAAGA	TAGGAAACC - AA	AAATGATAGG	GGGAATTGGA	GGTTTT
0	AGATACAGG TAGG	CCAAGATGA	TACAGTATTAG	- AGAAATGT	AATTTGCCAG	GAAGATGGTAA	ACC - AAAATG	ATACGGGGGA	ATTOGA
2	A <mark>gatacaggag</mark> ca	GATGATACA	GTATTAG - AGA	AATGAATTT	G C C A G G A A G A	TGGAAACC - AAA	AATGATAGGG	GGAATTGGAG	GTTTTA

## Issues with Multiple Alignments of short pyrosequencing reads

- Huge number of reads: ~100,000 reads in a single run
- Out of box alignment software are not feasible :
- Because of high complexity. Most are of the order of O(N<sup>4</sup>) e.g. Muscle, Clustalw
- Do not give accurate alignments for reads.
  - Primary reason: Do not take into account the position of the reads w.r.t. genome.

### **Pyro-Align Algorithm**

- Based on the parallel framework of Sample-Align-D algorithm
  - Sample-Align-D gives super-linear speed-ups on multiprocessors; hence gives speed advantages when used on sequential machines
- Assumptions:
  - The original genome or the wildtype is available
  - The reads are in 'forward' orientation

### **Pyro-Align Algorithm sketch**

- Align each read to the reference genome, using semi-global alignment.
  - This will place the reads in correct positions

#### >Do a Hierarchical Progressive alignment:

- Reorder the reads using the starting positions of the reads to 'generate' a guidance tree.
- >Do pairwise alignments according to the tree.
- >Do profile-profile alignments in a hierarchical fashion

#### Hierarchical Progressive alignment

Pairwise alignment



#### Pyro-Align Algorithm Complexity Analysis

- Semi-global alignment = O(N L<sub>R</sub> L<sub>G</sub>)
- Clustering & Reordering =  $O(N+NL_G)$
- Pairwisealignments  $= O(N L_R^2)$
- Profile-Profile alignments =  $O(N \log N L^2_G)$
- Total Asymptotic Complexity= O(N log N  $L^{2}_{G}$  + N  $L^{2}_{R}$ )
- Where:
- N= number of reads
- L<sub>R</sub> = Average length of reads
  - I \_I /I //I

#### Quality Assessment

- Objective:
  - To access the quality of the alignments w.r.t the original genome
  - Ensure that the system is able to handle reads from multiple haplotypes

#### • Problems:

- Ground truth for the alignments is not available(cannot be done with eyeballing due to large number of reads)
- Standard benchmarks such as Balibase or Prefab cannot be used.

#### Quality Assessment

- We choose HIV pol gene with length of 1970bp as wildtype for the experiments
- Four sets of genomes from the wildtype are produced at 3%, 5%, 7% and 10% mutations.
- These mutated genomes are used to produce the reads.
- The reads are aligned and the consensus from the reads is compared with the reference genome.

![](_page_14_Figure_0.jpeg)

Genomes obtained from different mutations of the wildtype

![](_page_14_Figure_2.jpeg)

The consensus obtained from the alignment is compared with the mutated genomes

#### Results

#### Successful alignment of reads from mutated genome reads as well as the mixture.

read_0863 beg 189 length 250/1-250	T-GG-A-C-AT	- AAA - G - C - T - A	- TA- GG-	TA-G-A-	G - T - A -	TTA-GTA-	GG - <mark>A</mark> - CC - 1	A - C - A - C C	- T - G - T -	C - AA - A - C	A -
read_0552 beg 187 length 250/1-250	T-GG-A-C-AT	· AAA · G · C · T · A	- TA - GG -	TACG-A-	G - T - A -	TTA-GTA-	G G - <mark>A</mark> - C C - 1	A - C - A - C C	- T - G - T -	C - AA - A	A .
read_0513 beg 191 length 251/1-251	T-GG-A-C-AT	· AAA · G · C · T · A	· TA · GG ·	TA-G-A-	G . <u>.</u> . <mark>A</mark> .	TTA-GTA-	GG · ACCC · 1	A · C · A · C C	. T . G . T .	C - AA - A - C	A -
read_1463 beg 192 length 250/1-250	T-GG-A-C-AT	· · A A · <mark>G</mark> · C · T · A	· TA · GG ·	TA-G-A-	G - T - A -	TTA · GTA ·	GG · A · CC · 🕇	A · C · A · C C	. T . G . T .	C - AA - A - C	A -
read_0642 beg 192 length 252/1-252	T-GG-A-C-AT	· AAA · <mark>G</mark> · C · T · A	• <b>T</b> A • G G •	TA-G-A-	G - T - A -	TTA-GTAG	GG · A · CC · 1	A · C · A · C C	. T . G . T .	C - AA - A - C	A -
read_1683 beg 192 length 253/1-253	T-GG-A-CGAT	· AAA · <mark>G</mark> · C · T · A	• T A • G G •	TA-G-A-	G - T - A -	ΤΤ	GG - <mark>A</mark> - CC - 1	A · C · A · C C	. T . G . T .	C - AA - A - C	A -
read_0720 beg 196 length 248/1-248	T - GG - A - C - A T	· AAA · <mark>G</mark> · C · T · A	· TA · GG ·	TA-G-A-	G - T - A -	TT··GTA·	GG · A · CC · 1	A · C · A · C C	. T . G . T .	C - AA - A - C	A -
read_0263 beg 197 length 249/1-249	TAGG - A - C - AC - T	· AAA · <mark>G</mark> · C · T · A	- TA - GG -	TA-G-A-	G - T - A -	TTA-GTA-	G G - <mark>A</mark> - C C - 1	A · C · A · C C	. T . G . T .	C - AA - A - C	A -
read_0320 beg 197 length 251/1-251	T-GG-A-C-AT	• <mark>AAA • <u>•</u> • C • T • A</mark>	- TA-GG-	TA-G-A-	G - T - A -	TTA-GTA-	G G - <mark>A</mark> - C C - 1	A - C - A - C C	- T - G - T -	C-AA-A-C	<mark>A</mark> -
read_0844 beg 197 length 255/1-255	T-GG-A-C-AC-T	· AAA · <mark>G</mark> · C · T · A	- TA-GG-	TA-G-A-	G - T - A -	TTA - GTA -	G G - <mark>A</mark> - C C - 1	A - C - A - C C	- T - G - T -	CGAA-A-C	A -
read_0349 beg 197 length 252/1-252	TGGG-A-C-AT	· AAA · <mark>G</mark> · C · T · A	- <mark>T</mark> A - <mark>GG</mark> -	TA-G-AC	G - T - A -	TTA-GTA-	66 - <mark>A</mark> - CC - 1	A - C - A - C C	• T • G • T •	C - AA - A - C	<mark>A</mark> -
read_1694 beg 198 length 252/1-252	T-GG-A-C-AT	· AAA · G · C · T · A	- TA-GG-	TA-GTA-	G - T - A -	TTA - GTA -	66 · <mark>A</mark> · C C · 🚺	A - C - A - C C	. T . G . T .	AA - A - C	<mark>A</mark> -
read_1188 beg 198 length 253/1-253	T-GG-A-C-AC-T	· AAA · <mark>G · C · T ·</mark> A	- <mark>T A</mark> - <mark>G G</mark> -	T A - G - A -	G - T - A -	TTA-GTA-	GG - <mark>A</mark> - CC - 1	A - C - A - C C	· T · G · T ·	C - AA - A - C	A -
read_0549 beg 203 length 256/1-256	· · · <mark>G G</mark> · <mark>A</mark> · C · <mark>A</mark> · · T	· AAA · <mark>G T C · T · A</mark>	• TA • GG •	TA-G-A-	G - T - A -	TTA · GTA · ·	66 · <mark>A</mark> · CC · 1	A · C · A · C C	· T · G · T ·	C - AA - A - C	A -
read_1818 beg 203 length 252/1-252	<mark>G G</mark> - <mark>A - C - A</mark> T	· AAA · G · C · T · A	• T A • • G •	TACG-A-	G - T - A -	TTA - GTA -	GG · <mark>A</mark> · CC · 1	A - CGA - CC	· T · G · T ·	C - AA - A - C	A -
read_0370 beg 204 length 249/1-249	· · · G · A · C · A · · T	· AAA · G · C · T · A	• <mark>T A</mark> • G G •	TA-G-A-	G - T - A -	TTA-GTA-	GG · A · CC ·	A · C · A · C C	· T · G · T ·	C - AA - A - C	<mark>A</mark> -
read_0596 beg 205 length 250/1-250	<mark>A . C</mark> . A T	· AAA · <mark>G · CG T</mark> · A	• <b>T</b> A • GG •	TA-G-A-	G - T - A -	TTA · GTA ·	GG · A · CC · 1	A · C · A · C C	· T · G · T ·	C - AA - A - C	A -
read_1416 beg 205 length 255/1-255	· · · · · <mark>A · C · A</mark> · · T	· AAA · <mark>G · C · T ·</mark> A	· TA · GG ·	TA-G-A-	G · T · A ·	TTA · GTA · I	GGTA - CCG	A · C · A · C C	· TAG · T ·	C - AA - A - C	A -
read_0574 beg 206 length 253/1-253	· · · · · · · C · A · · T	· AAA · <mark>G · C · T ·</mark> A	· TA · GG ·	TA-G-A-	G · T · A ·	TTA - GTA -	GG · A · CC · T	A · C · A · C C	· T · G · T ·	C A - A - C	A -
read_0087 beg 207 length 254/1-254	<mark>A</mark> T	· AAA · <mark>G</mark> · C · T · A	• TA • GG •	TACG - A -	G - T - A -	TTA · GTA · ·	GG · A · CC ·	A · C · A · C C	· T · G · T ·	C - AA - A - C	<mark>A</mark> -
read_1256 beg 207 length 251/1-251	· · · · · · · · · A · · T	• AAA • <mark>G • C • T •</mark> A	- TA - GG -	TA·G·A·	G - T - A -	TTA · G · A ·	GG · A · CC ·	A · C · A · CC	· TAG · T ·	C - AA - A - C	A -
read_1328 beg 208 length 259/1-259		• AAA • <mark>G • C • T •</mark> A	- TA - GG -	TAIGIAI	G - T - A -	TTACGTA-	66 · <mark>A</mark> · CC · 🛛	A · C · A · C C	· T · G · T ·	C - AA - A - C	A -
read_1274 beg 209 length 252/1-252		• AAA • <mark>6</mark> • C • T • A	· TA · GG ·	TA-G-A-	G · T · A ·	TTA · GTA ·	GG · <mark>A</mark> · CC · ·	A · C · A · CC	· T · G · T ·	C - AA - A - C	A -
read_0560 beg 209 length 254/1-254		• 🗛 🗛 • 💁 • 🕻 • T • A	- TA- GG-	TA-G-A-	G - T - AC	TTA - GTA -	G G - <mark>A</mark> - C C - 1	A · C · A · CC	· T · G · T ·	C - AA - A - C	A -
read_1426 beg 213 length 252/1-252		· · · · · · · <mark>C</mark> · <u>T</u> · A	• TA • GG •	TA-G-A-	G - T - A -	TTA · GTA ·	GG · A · CC ·	A - C - A - C C	· T · G · T ·	C - AA - A - C	A -
read_0893 beg 215 /ength 250/1-250		<mark>A</mark>	- TA- GG-	TA-G-A-	G - T - A -	TTA-GTA-	G G - A - C C - T	A - C - A - C C	· T · G · T ·	C - AA - A - C	A -
read_1490 beg 216 length 258/1-258			- TA- GG-	TA-G-A-	G - T - A -	TTA - GTA -	G G - <mark>A</mark> - C C - T	A - C - A - C C	· T · G · T ·	C - AA - A - C	A -
read_0547 beg 216 length 254/1-254			• TA • GG •	TA - G - A -	G - T - A -	TTA - GTA -	G G - <mark>A</mark> - C C - T	A - C - A - C C	· T · G · T ·	C - AA - A - C	· · A ·
read_0069 beg 218 length 254/1-254			· · · · GG ·	TA - G - A -	G - T - A -	TTA-GTA-	G G - A - C C - T	A - C - A - C C	· T · G · T ·	C - AA - A - C	A .
read_0028 beg 218 /ength 254/1-254			<mark>GG</mark> .	TA · G · A ·	G - T - A -	TTA-GTA-	G G - <mark>A</mark> - C C - 1	A - C - A - C C	· T · G · T ·	C - AA - A - C	· · A ·
read_0116 beg 220 length 256/1-256				TA · G · A ·	G - T - A -	TTA - GTA -	G G - A - C C - T	A - C - A - C C	· T · G · T ·	C - AA - A - C	A .
read_0613 beg 222 length 257/1-257				••• <mark>•</mark> • <mark>•</mark> •	G - T - A -	TTA-GTA-	GG - A - CCT	A - C - A - C C	· T · G · T ·	C - AA - A - C	· · A ·
read_0382 beg 222 length 264/1-264				· · · G · A ·	G . T . A .	TTA-GTA-	GG - A - CC - 1	A - C - A - C C	· T · 📴 · T ·	C - AA - A - C	- <mark>G</mark> A -
read_0340 beg 222 length 249/1-249				• • • 📴 • 🗛 •	🖸 · T · A ·	TTA-GTA-	GG · A · CC ·	A - C - A - C C	· T · G · T ·	C - AA - A - C	· · A ·
read_0454 beg 225 length 251/1-251					· · T · A ·	TTA-GTA-	GG · A · CC ·	A - C - A - C C	· • • • • • •	C - AA - A - C	· · A ·
read 1909 beg 225 length 253/1-253					T - A -	TTA · GTA ·	GG - 🗛 - CC - 🚺	A - CGA - CC	- T - G - T -	C - AA - A - C	A -

![](_page_16_Picture_0.jpeg)

![](_page_16_Picture_1.jpeg)

![](_page_17_Picture_0.jpeg)

![](_page_17_Picture_1.jpeg)

![](_page_18_Picture_0.jpeg)

![](_page_18_Picture_1.jpeg)

## Ongoing Works—Sketch of parallelization of pyro-Align

- Proposed Approach:
- In parallel on multiple processors, align each read to the reference genome, using semi-global alignment.
- Do a Parallel Hierarchical Progressive alignment using sample-align-D
- We expect super-linear speedups for parallel pyro-align giving enormous advantage in terms of timing and memory

#### **Conclusions and Future work**

- A low complexity algorithm for aligning huge number of pyrosequencing reads is presented.
- Successfully aligned the reads from mutated and mixture of mutated genomes.
- Presented the quality assessment and compared with pairwise alignments
- We are working on parallelization of the algorithm

# Acknowledgements

- We would like to thank the Beerenwinkel group for the support at Biosystems science and engineering dept at ETH Zurich.
- Thanks also goes to Nick Eriksson of the University of Chicago
- Special thanks to Tanya Berger Wolf of Laboratory for Computational Population Biology, UIC.

![](_page_22_Picture_0.jpeg)

![](_page_22_Picture_1.jpeg)