

# 생명정보학에서 쓰이는 컴퓨터 Perl 언어의 기초 교육

Raw sequences of NGS technologies and Perl program

2017/9/23

Jongsun Park, Ph. D.



서울시 강서구 화곡동 359-63 영주빌딩 201호

Tel 02 2698 1188 / Fax 02 6280 8821

[www.infoboss.co.kr](http://www.infoboss.co.kr)

**InfoBoss**  
Informations & Systems

- Two major rice genomes were already published in 2002.

## A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. *japonica*)

Stephen A. Goff,<sup>1\*</sup> Darrell Ricke,<sup>1</sup> Tien-Hung Lan,<sup>1</sup>  
 Gernot Presting,<sup>1</sup> Ronglin Wang,<sup>1</sup> Molly Dunn,<sup>1</sup>  
 Jane Glazebrook,<sup>1</sup> Allen Sessions,<sup>1</sup> Paul Oeller,<sup>1</sup> Hemant Varma,<sup>1</sup>  
 David Hadley,<sup>1</sup> Don Hutchison,<sup>1</sup> Chris Martin,<sup>1</sup> Fumiaki Katagiri,<sup>1</sup>  
 B. Markus Lange,<sup>1</sup> Todd Moughamer,<sup>1</sup> Yu Xia,<sup>1</sup> Paul Budworth,<sup>1</sup>  
 Jingping Zhong,<sup>1</sup> Trini Miguel,<sup>1</sup> Uta Paszkowski,<sup>1</sup> Shiping Zhang,<sup>1</sup>  
 Michelle Colbert,<sup>1</sup> Wei-lin Sun,<sup>1</sup> Lili Chen,<sup>1</sup> Bret Cooper,<sup>1</sup>  
 Sylvia Park,<sup>1</sup> Todd Charles Wood,<sup>2</sup> Long Mao,<sup>3</sup> Peter Quail,<sup>4</sup>  
 Rod Wing,<sup>5</sup> Ralph Dean,<sup>5</sup> Yeisoo Yu,<sup>5</sup> Andrey Zharkikh,<sup>6</sup>  
 Richard Shen,<sup>6†</sup> Sudhir Sahasrabudhe,<sup>6</sup> Alun Thomas,<sup>6</sup>  
 Rob Cannings,<sup>6</sup> Alexander Gutin,<sup>6</sup> Dmitry Pruss,<sup>6</sup> Julia Reid,<sup>6</sup>  
 Sean Tavtigian,<sup>6</sup> Jeff Mitchell,<sup>6</sup> Glenn Eldredge,<sup>6</sup> Terri Scholl,<sup>6</sup>  
 Rose Mary Miller,<sup>6</sup> Satish Bhatnagar,<sup>6</sup> Nils Adey,<sup>6</sup>  
 Todd Rubano,<sup>6†</sup> Nadeem Tusneem,<sup>6</sup> Rosann Robinson,<sup>6</sup>  
 Jane Feldhaus,<sup>6</sup> Teresita Macalma,<sup>6</sup> Arnold Oliphant,<sup>6†</sup>  
 Steven Briggs<sup>1</sup>

## A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. *indica*)

Jun Yu,<sup>1,2,3,4\*</sup> Songnian Hu,<sup>1\*</sup> Jun Wang,<sup>1,2,5\*</sup>  
 Gane Ka-Shu Wong,<sup>1,2,4\*</sup> Songgang Li,<sup>1,5</sup> Bin Liu,<sup>1</sup> Yajun Deng,<sup>1,6</sup>  
 Li Dai,<sup>1</sup> Yan Zhou,<sup>2,7</sup> Xiuqing Zhang,<sup>1,3</sup> Mengliang Cao,<sup>8</sup> Jing Liu,<sup>2</sup>  
 Jiandong Sun,<sup>1</sup> Jiabin Tang,<sup>1,3</sup> Yanjiong Chen,<sup>1,6</sup>  
 Xiaobing Huang,<sup>1</sup> Wei Lin,<sup>2</sup> Chen Ye,<sup>1</sup> Wei Tong,<sup>1</sup> Lijuan Cong,<sup>1</sup>  
 Jianing Geng,<sup>1</sup> Yujun Han,<sup>1</sup> Lin Li,<sup>1</sup> Wei Li,<sup>1,9</sup> Guangqiang Hu,<sup>1</sup>  
 Xiangang Huang,<sup>1</sup> Wenjie Li,<sup>1</sup> Jian Li,<sup>1</sup> Zhanwei Liu,<sup>1</sup> Long Li,<sup>1</sup>  
 Jianping Liu,<sup>1</sup> Qiuhui Qi,<sup>1</sup> Jinsong Liu,<sup>1</sup> Li Li,<sup>1</sup> Tao Li,<sup>1</sup>  
 Xuegang Wang,<sup>1</sup> Hong Lu,<sup>1</sup> Tingting Wu,<sup>1</sup> Miao Zhu,<sup>1</sup>  
 Peixiang Ni,<sup>1</sup> Hua Han,<sup>1</sup> Wei Dong,<sup>1,3</sup> Xiaoyu Ren,<sup>1</sup>  
 Xiaoli Feng,<sup>1,3</sup> Peng Cui,<sup>1</sup> Xianran Li,<sup>1</sup> Hao Wang,<sup>1</sup> Xin Xu,<sup>1</sup>  
 Wenxue Zhai,<sup>3</sup> Zhao Xu,<sup>1</sup> Jinsong Zhang,<sup>3</sup> Sijie He,<sup>3</sup>  
 Jianguo Zhang,<sup>1</sup> Jichen Xu,<sup>3</sup> Kunlin Zhang,<sup>1,5</sup> Xianwu Zheng,<sup>3</sup>  
 Jianhai Dong,<sup>2</sup> Wanyong Zeng,<sup>3</sup> Lin Tao,<sup>2</sup> Jia Ye,<sup>2</sup> Jun Tan,<sup>2</sup>  
 Xide Ren,<sup>1</sup> Xuwei Chen,<sup>3</sup> Jun He,<sup>2</sup> Daofeng Liu,<sup>3</sup> Wei Tian,<sup>2,6</sup>  
 Chaoguang Tian,<sup>1</sup> Hongai Xia,<sup>1</sup> Qiyu Bao,<sup>1</sup> Gang Li,<sup>1</sup> Hui Gao,<sup>1</sup>  
 Ting Cao,<sup>1</sup> Juan Wang,<sup>1</sup> Wenming Zhao,<sup>1</sup> Ping Li,<sup>3</sup> Wei Chen,<sup>1</sup>  
 Xudong Wang,<sup>3</sup> Yong Zhang,<sup>1,5</sup> Jianfei Hu,<sup>1,5</sup> Jing Wang,<sup>1,5</sup>  
 Song Liu,<sup>1</sup> Jian Yang,<sup>1</sup> Guangyu Zhang,<sup>1</sup> Yuqing Xiong,<sup>1</sup> Zhijie Li,<sup>1</sup>  
 Long Mao,<sup>3</sup> Chengshu Zhou,<sup>8</sup> Zhen Zhu,<sup>3</sup> Runsheng Chen,<sup>1,9</sup>  
 Bailin Hao,<sup>2,10</sup> Weimou Zheng,<sup>1,10</sup> Shouyi Chen,<sup>3</sup> Wei Guo,<sup>11</sup>  
 Guojie Li,<sup>12</sup> Siqi Liu,<sup>1,2</sup> Ming Tao,<sup>1,2</sup> Jian Wang,<sup>1,2</sup> Lihuang Zhu,<sup>3†</sup>  
 g<sup>1,2,3†</sup>

The genome of the japonica subspecies of rice, an important cereal and model monocot, was sequenced and assembled. The assembled sequence was used for gene predictions on the assembly. Between 32,000 to 50,000 genes were predicted. Some barley proteins are found in the other cereal genomes; however, the number is limited. Assignment of candidates to genes is limited in many cases. The rice genome is a valuable resource for improvement of cereals, our

	<i>Oryza sativa</i> subsp. <i>japonica</i>	<i>Oryza sativa</i> subsp. <i>indica</i>
Genome length	374,471,240	395,818,884
# of scaffolds	12	2,946
GC ratio (%)	43.158	43.569
# of genes	55,986	37,358
# of transcripts/proteins	66,338	37,358

is the most widely distributed rice subspecies, by whole-genome size, with an estimated genome size of 395.8 Mb. The assembled sequences were used for gene predictions on the assembly. Between 32,000 to 50,000 genes were predicted. Some barley proteins are found in the other cereal genomes; however, the number is limited. Assignment of candidates to genes is limited in many cases. The rice genome is a valuable resource for improvement of cereals, our

- Running program for aligning short reads against the reference genome.

## Alignment process with bwa (for each fastq file)

```
/home/bioinfo/bwa-0.7.15/bwa aln -n 3 -t 1 -o 1 -e -1 -d 16 -i 5 -l 45 -k 3 -M 3 -O 11 -E 4  
/home/bioinfo/bwadatastet//TBBWA0008776.fasta [filename] > [filename].sai  
/home/bioinfo/bwa-0.7.15/bwa aln -n 3 -t 1 -o 1 -e -1 -d 16 -i 5 -l 45 -k 3 -M 3 -O 11 -E 4  
/home/bioinfo/bwadatastet//TBBWA0008776.fasta [filename] > [filename].sai
```

## Merging alignment results and generate bam file

```
/home/bioinfo/bwa-0.7.15/bwa sampe /home/bioinfo/bwadatastet//TBBWA0008776.fasta [sai  
file1] [sai file2] [filename1] [filename2] -r  
"@RG\tID:SR15950\tPL:illumina\tPU:seq\tSM:illumina\tLB:SR" | /home/bioinfo/samtools-  
1.5//samtools view -bt /home/bioinfo/bwadatastet//TBBWA0008776.txt -o [bam filename] -
```

- Getting statistics of alignment from bam file

Get flagstat data from bam

```
/home/bioinfo/samtools-1.5//samtools flagstat [bam filename]
```

PC32.pl

- Merge two flagstat files via extracting numbers from text file and print it.

*Thank you for your attention!*

*If you have question,  
please ask! =)*

**starflr@infoboss.co.kr**

**InfoBoss**  
Informations & Systems