

## Biofuel과 Omics Transcriptomics에 대해서

박정진

지금 현재, 전세계에서 biofuel연구를 가장 활발히 지원하고 있는 나라는 미국일 것이다. 미국 에너지성 (DOE) 뿐만 아니라, 공군 그리고 민간 단체까지 합하면 일년에 들이는 연구비만 1500억 정도인 것으로 알려져 있다. 2010년 DOE가 지원하는 컨소시엄중 하나인 NAABB (National Alliance for Advanced Biofuels and Bioproducts, <http://www.naabb.org>)의 경우 작년 한해 동안만 미화 4천 4백만불을 연구비로 사용했다. 이런 막대한 금액의 연구비는 다양한 분야의 연구 활동을 지원하게 되는데, 그 중 하나가 바로 transcriptomic analysis이다. 현재 DNA 서열정보와 DNA microarray나 deep sequencing은 genome 연구에 있어서 중요한 도구가 되었으며, 유전자 발현 양상을 대량으로 분석하는데 주로 이용되고 있다. 한 생물 종의 전체 genome으로부터 발현되는 전체 유전체를 총괄하여 transcriptome이라고 부르며, 이 transcriptome의 발현 패턴을 알아보는 것을 transcriptome 분석이라고 한다.

지금까지 개발된 다양한 종류의 유전자 변형 방법은 주로 모델 시스템인 대장균, 효모와 같은 단세포 생물이나 애기장대와 같이 제한된 종류의 개체에만 적용될 수 있었다. 왜냐하면 유전자 변형의 기본이라 할 수 있는 genome sequence를 알기 위해서는 막대한 자금과 시간이 필요했기 때문이다. 하지만 454 sequencing을 필두로, illumine sequencing과 같은 차세대 sequencing 기술이 개발됨에 따라 그 비용과 시간을 획기적으로 줄일 수 있게 되었다.

## Every genome tells a story. What's yours?



Uncover your story

- How can changes in your DNA impact your health?
- What is a genetic disease and how is it passed from one generation to the next?

### Access the bigger picture for your patients

#### Through comprehensive genetic sequence information.

- Individual genome sequencing: \$19,500 per genome
- Medically indicated cases approved through the Illumina subsidy program: \$9,500 per genome
- Groups of five individuals ordered through the same physician: \$14,500 per genome

#### Contact Us

First Name: \*

Last Name: \*

<그림 1> Illumina에서 지원하는 everygenome 서비스 웹사이트 (<http://www.everygenome.com/>)

<그림1>은 Illumina에서 현재 진행중인 everygenome 서비스라는 personal genome sequencing 서비스이다. 웹사이트의 대문에 걸려있는 내용을 보면, 개인의 genome sequence를 미화 1만 9천 5백불에 서비스 하고 있다. 즉 2천만원 조금 넘는 돈으로 자신의 genome sequence를 알 수 있다는 이야기다.

이렇게 단가가 낮아진 관계로, 요즘 biofuel이나 의약품의 원료로 각광을 받고 있는 식물들에 대해서 sequencing을 하는 것이 유행처럼 번지게 되었다.



## The 1KP Project

### Links

[Home](#)

[What is the 1KP Project?](#)

[Why Sequence 1000 plants?](#)

[Transcriptomes not Genomes](#)

[Essential Plant Phylogeny](#)

[Media](#)

[Contact Us](#)

The 1000 plants (oneKP or 1KP) initiative is a public-private partnership generating large scale gene sequence information for 1000 different species of plants. Major supporters include Alberta's Department of Advanced Education and Technology (AET), Silicon Valley based Musea Ventures, Beijing Genomics Institute in Shenzhen, University of Alberta, and Alberta's Informatics Circle of Research Excellence (iCORE).

## Subproject Categories

### Grouped by Phylogeny

[Angiosperms](#)

[Non Flowering](#)

[Green Algae](#)

### Grouped by Application

[Agricultural](#)

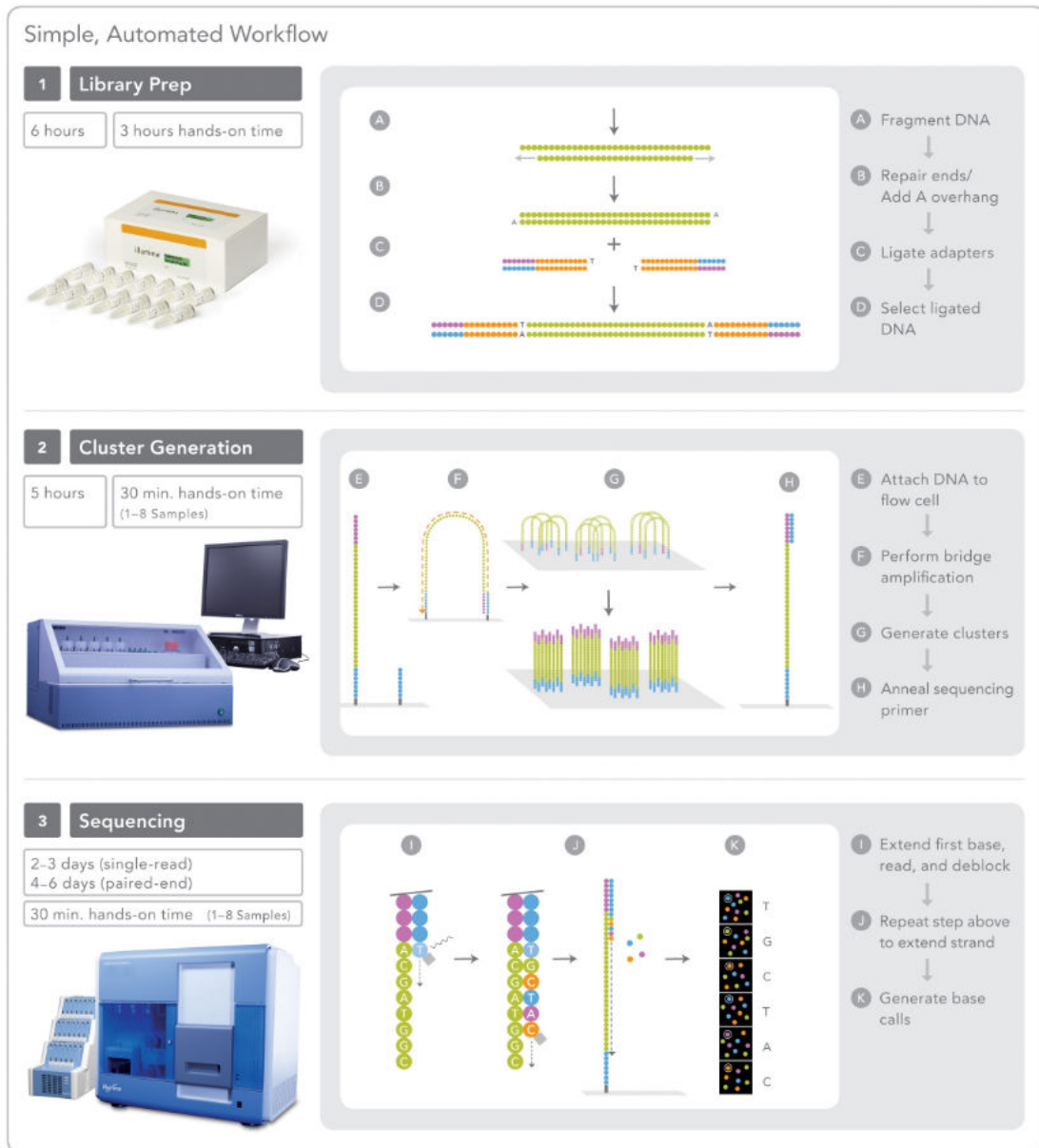


<그림 2> 캐나다, 미국, 중국이 진행하고 있는 1000 plant transcriptome 프로젝트 ([www.onekp.com](http://www.onekp.com))

<그림 2>는 바로 transcriptomics 연구중 하나인 1KP 프로젝트 웹사이트 대문으로, 천여 종의 녹조류와 식물에 대해 transcriptome analysis를 2008년부터 진행해 오고 있다. 이러한 프로젝트가 진행될 수 있었던 가장 큰 이유는, 앞서서도 말한 차세대 sequencing 덕분이라고 할 수 있다.

<그림 3>은 차세대 sequencing 중 하나인 Illumina sequencing의 작업 단계를 나타내고 있다. 우선 1에서 10 µg의 total RNA를 가지고, Dynabeads의 mRNA isolation kit으로 mRNA를 분리해 낸다. 분리해 낸 mRNA를 fragmentation한 후, sequencing adapter를 붙이고, agarose gel에서 적절한 사이즈의 band를 떼어낸다. 그 후, PCR을 하면 일차적으로 cDNA library제작이 끝나게 된다. 이런 library제작용 kit의 경우 시약 제조 회사에 따라 미화 500불에서 3000불 (library 10개 제작용) 정도 한다. 이렇게 만든 library를 flow cell에 올려 clustering을 만들고, 만들어진 flow cell을 GAIIx나 HiSeq 2000과 같은 sequencer에 넣어 sequencing 데이터를 얻게 된다. 하지만 Illumina sequencing의 경우 한번에 읽을 수 있는 길이가 50bp 내외여서 data assemble이 data의 quality를 좌우한다고 할 수 있다. 일차 data 분석이 끝나면 보통 5만개에서 10만개의 contig에 대한 정보 (sequence, read count number)를 얻게 된다. 기존의 microarray에 대해 장점으로는 직접 sequencing을 하기 때문에 full genome sequence가 안 알려져 있어도 사용할 수 있으며 가격이 저렴하다는 것과 발현량 측정에 한계가 없다는 것을 들 수 있다(일반적으로

microarray는 16비트 방식의 스캐너를 사용하게 되어 0에서 65535 사이의 값을 갖게 된다). 하지만 assembling algorithm에 따라 결과가 바뀔 수 있으며, PCR 과정을 거치는 동안 발현량이 왜곡될 수 있다는 단점이 있다.



©2008, Illumina Inc. All rights reserved.

<그림 3> Illumina sequencing work flow

<표 1> Bioenergy 생산에 이용되는 생물 종의 genome 정보 (Rubin, E. M. (2008). "Genomics of cellulosic biofuels." Nature 454(7206): 841-845)

**Table 1 | Bioenergy genomes**

Organism	Genome size (megabases)	Status	Reference
<b>Feedstocks and feedstock models</b>			
<i>Populus trichocarpa</i> (poplar)	480	Complete	Ref. 9
<i>Chlamydomonas reinhardtii</i>	120	Complete	Ref. 34
<i>Glycine max</i> (soya bean)	1,200	Draft	-
<i>Manihot esculenta</i> (cassava)	770	In progress	-
<i>Sorghum bicolor</i>	760	In progress	-
<i>Eucalyptus globulus</i>	600	In progress	-
<i>Brachypodium distachyon</i>	355	In progress	-
<i>Zea mays</i> (maize)	2,500	In progress	-
<i>Elaeis guineensis</i> (oil palm)	~3,400	In progress	<a href="http://www.checkbiotech.org/green_News_Biofuels.aspx?infoId=15100">http://www.checkbiotech.org/green_News_Biofuels.aspx?infoId=15100</a>
<i>Panicum virgatum</i> (switchgrass)	~5,600	In progress	-
<i>Setaria italica</i> (foxtail millet)	~515	In progress	-
<b>Biomass degraders</b>			
<i>Acidothermus cellulolyticus</i> 11B	2.4	Complete	-
<i>Bacillus pumilus</i> SAFR-032	3.7	Complete	Ref. 35
<i>Caldicellulosiruptor saccharolyticus</i> DSM 8903	3.0	Complete	-
<i>Clostridium phytofermentans</i> ISDg	4.8	Complete	-
<i>Clostridium thermoceillum</i> ATCC 27405	3.8	Complete	-
<i>Cytophaga hutchinsonii</i> ATCC 33406	4.4	Complete	-
<i>Flavobacterium johnsoniae</i> UW101	6.1	Complete	-
<i>Rubrobacter xylophilus</i> DSM9941	3.2	Complete	-
<i>Saccharophagus degradans</i>	5.1	Complete	Ref. 36
<i>Thermobifida fusca</i> strain YX	3.6	Complete	Ref. 37
<i>Clostridium cellulolyticum</i> H10	4.0	Draft	-
<i>Elusimicrobium minutum</i> PeU91	1.6	Draft	-
<i>Nectria haematococca/Fusarium solani</i>	51	Draft	-
<i>Phanerochaete chrysosporium</i>	35.1	Draft	-
<i>Pastia placenta</i>	33	Draft	-
<i>Sagittula stellata</i> E-37	5.3	Draft	-
<i>Trichoderma reesei/Hypocrea jecorina</i>	33	Draft	-
<i>Cellulomonas flavigena</i> DSM 20109	~4.0	In progress	-
<i>Cellvibrio japonicus</i> Ueda107	~6.0	In progress	-
<i>Fibrobacter succinogenes</i> subsp. <i>succinogenes</i> S85	~3.8	In progress	-
<i>Ruminococcus albus</i>	4.0	In progress	-
<i>Teredinibacter turnerae</i> T7902	~2	In progress	-
Termite hindgut community	NA	Complete	Ref. 23
Poplar biomass degrading community	NA	In progress	<a href="http://www.jgi.doe.gov/sequencing/ispaseqplans2007.html">http://www.jgi.doe.gov/sequencing/ispaseqplans2007.html</a>
Asian longhorned beetle ( <i>Anoplophora glabripennis</i> ) gut community	NA	In progress	<a href="http://www.jgi.doe.gov/sequencing/DOEmicrobes2007.html">http://www.jgi.doe.gov/sequencing/DOEmicrobes2007.html</a>
Bovine rumen community transcriptome	NA	In progress	<a href="http://www.energybiosciencesinstitute.org/index.php?option=com_content&amp;task=view&amp;id=159&amp;Itemid=20">http://www.energybiosciencesinstitute.org/index.php?option=com_content&amp;task=view&amp;id=159&amp;Itemid=20</a>
<b>Fuel producers</b>			
<i>Clostridium acetobutylicum</i> ATCC 824	4.0	Complete	Ref. 38
<i>Clostridium beijerinckii</i> NCIMB 8052	6.0	Complete	-
<i>Pichia stipitis</i>	15.4	Complete	Ref. 27
<i>Thermoanaerobacter tengcongensis</i> MB4	2.7	Complete	Ref. 39
<i>Zymomonas mobilis</i> subsp. <i>mobilis</i> ZM4	2.1	Complete	Ref. 40
<i>Bacillus coagulans</i> 36D1	2.9	Draft	-
<i>Thermoanaerobacter pseudethanolicus</i> 39E	2.4	Draft	-
<i>Clostridium ljungdahlii</i>	~4.0	In progress	-

Bioenergy-relevant organisms for which large-scale genome projects have been completed or are under way are listed. Information on genome projects without references can be found at <http://www.ncbi.nlm.nih.gov/sites/entrez?db=genomeprj>.

<표 1>에서는 지금 현재 biofuel에 사용되는 생물종과 모델이 되는 생물종의 genome size를 나타내고 있다. 최근 바이오 에너지에 대한 수요와 기대치가 올라감에 따라 이미 대부분의 생물종 들에 대한 full genome sequencing이 끝난 상태이다.

이러한 방대한 정보에서 우리가 원하는 정보를 얻기 위해서는 transcriptomics 뿐만 아니라 다른 omics 데이터와 함께 분석할 필요가 있을 것이다.

다음 장에서는 proteomics에 대해 알아보도록 하겠다.