CHANDRA SHEKHAR AZAD UNIVERSITY OF AGRICULTURE AND TECHNOLOGY, KANPUR- 208 002

Ph.D. Semester I

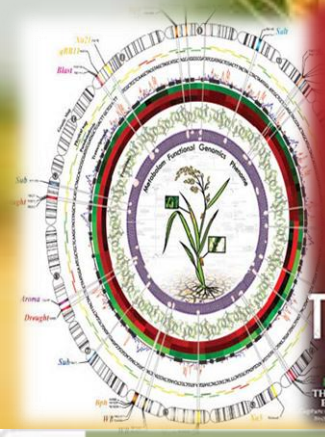
Course : Genomics in Plant Breeding (GPB-605)

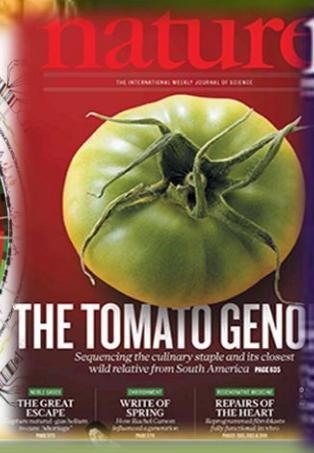
Lecture by: Sanjana Pathak

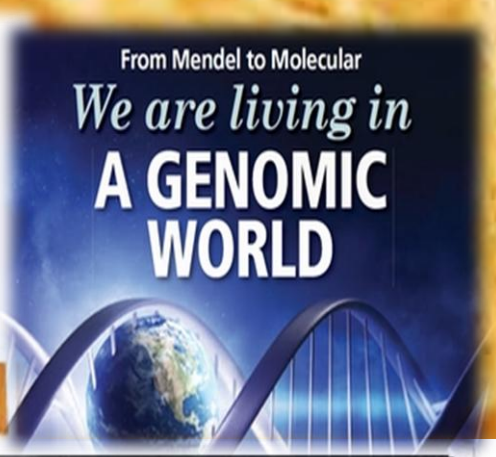
Teaching associate

Department of Genetics and Plant Breeding

ROLE OF FUNCTIONAL GENOMICS IN CROP IMPROVEMENT







WHAT IS GENOMICS?

- GENOME A genome can be defined as a complete set (n) of chromosomes (hence of genes) inherited as a unit from one parent. It is thus the entire genetic compliment of a living organism.
- GENOMICS is a discipline in genetics that applies recombinant DNA technology, DNA sequencing methods and bioinformatics to sequence, assemble and analyze the function and structure of genomes.

BRANCHES OF GENOMICS

STRUCTUAL GENOMICS

- Refers to high-throughput three dimensional structural determination and analysis of biological macromolecules.

FUNCTIONAL GENOMICS

- Development and application of global experimental approaches to assess gene functions by making use of the information and reagents provided by structural genomics.

COMPARATIVE GENOMICS

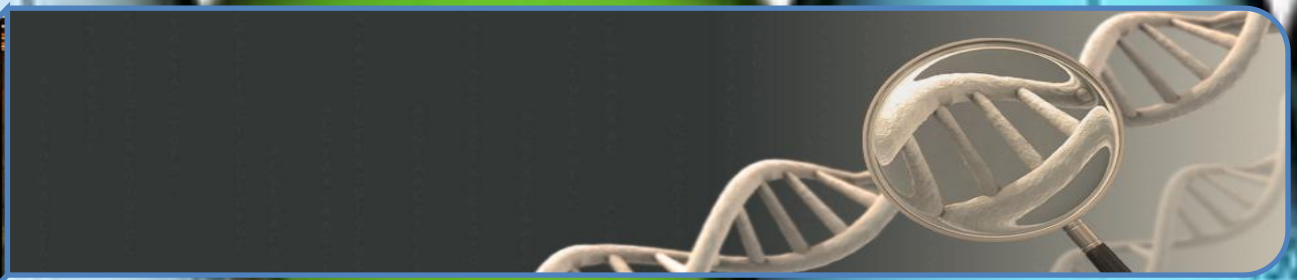
- Field of biological research in which the genomic features of different organisms are compared.

EVOLUTIONARY GENOMICS

- Deals with the study of evolution of genomes.

PHARMACO-GENOMICS

- Deals with the interaction of drugs with genomes.

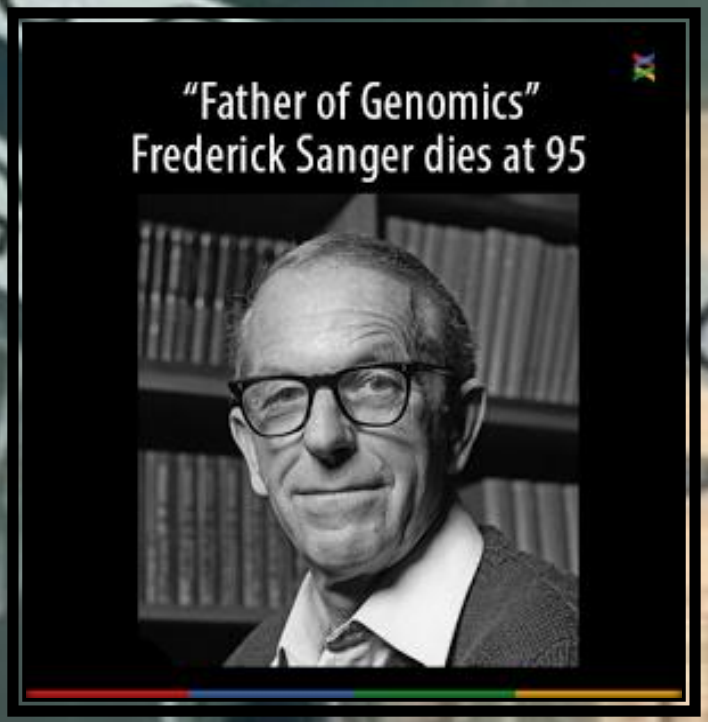


THE MAN WHO STARTED IT ALL

In 1977 Fredrick Sanger developed the classical DNA sequencing technique now known as the Sanger method, to determine the order of bases in a DNA strand.

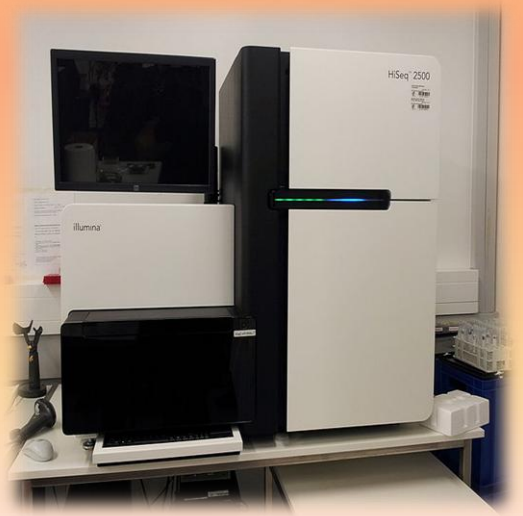
He sequenced the genome of the virus Bacteriophage ϕ×174.

It earned him his second Nobel Prize in Chemistry in 1980.



NEXT GENERATION GENOME SEQUENCING METHODS

- Massively Parallel Signature sequencing
- Polony sequencing
- 454 Pyrosequencing
- Illumina (Solexa) sequencing
- SOLiD sequencing
- Ion Torrent Semiconductor sequencing
- DNA Nanoball sequencing
- Heliscope single molecule sequencing
- Single Molecule Real Time sequencing



HIGH-THROUGHPUT GENOME SEQUENCING TECHNOLOGIES

454 / Roche

450 bp 1.5 Gbp / day

Illumina

150 bp 35 Gbp / day

Helicos

55 bp 4.5 Gb / day

SOLiD ABI

75 bp 22 Gbp / day

PacBio

1000 bp 70 Gbp / day

Ion PGM

100 bp 120 Gbp / day

GENOMICS CENTERS

wellcome trust sanger institute

BCM Baylor College of Medicine

BROAD INSTITUTE

NATIONAL INSTITUTE OF PLANT GENOME RESEARCH

JGI DOE JOINT GENOME INSTITUTE US DEPARTMENT OF ENERGY OFFICE OF SCIENCE

J. Craig Venter INSTITUTE

THE Genome CENTER AT WASHINGTON UNIVERSITY

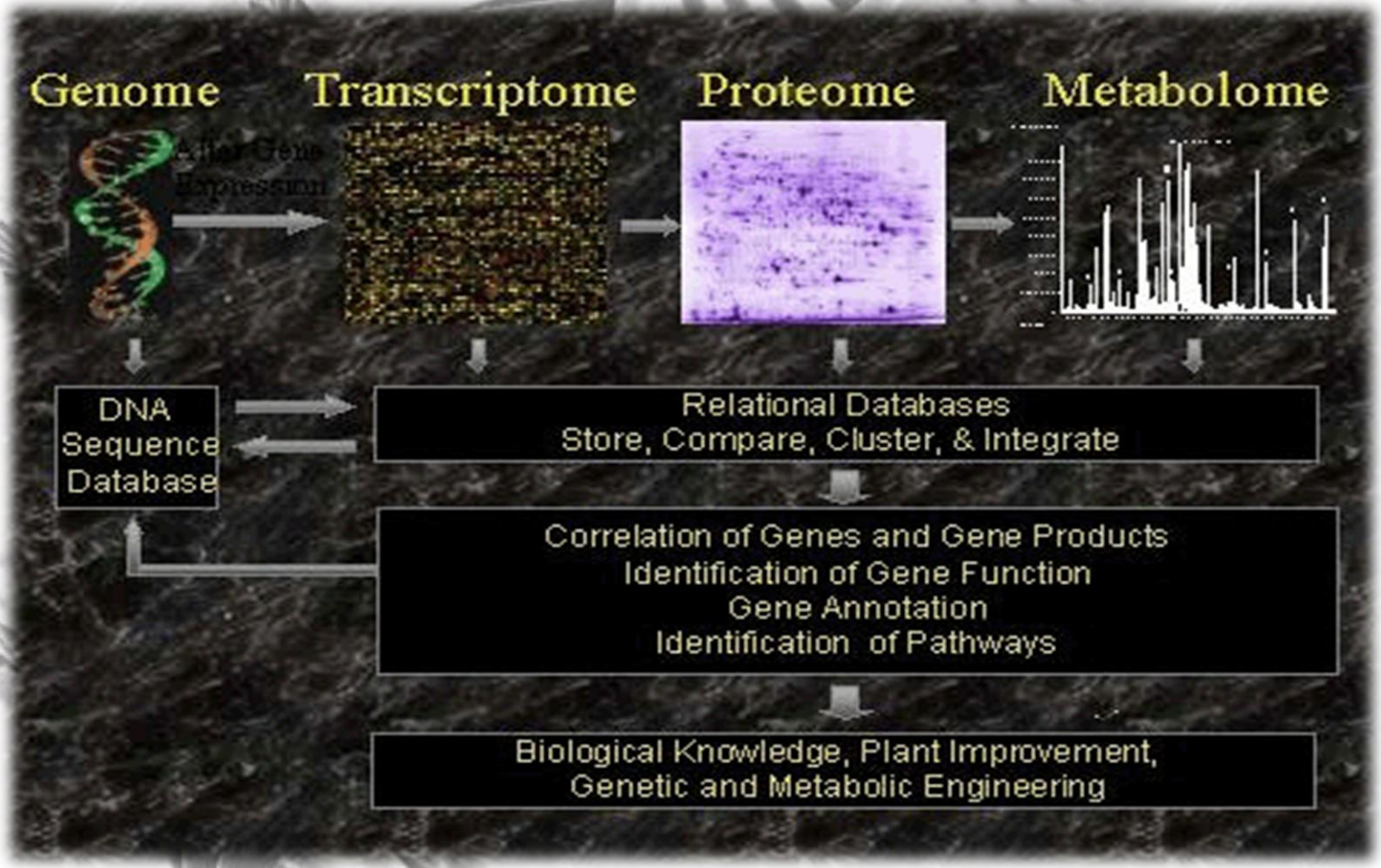
NATIONAL CENTRE FOR PLANT GENOME RESEARCH
NCPGR

NIBMG NATIONAL INSTITUTE OF BIOMEDICAL GENOMICS

Accelerating Genomics for Health

The ‘Omics’ Era

GENOMICS **TRANSCRIPTOMICS** **PROTEOMICS** **METABOLOMICS**

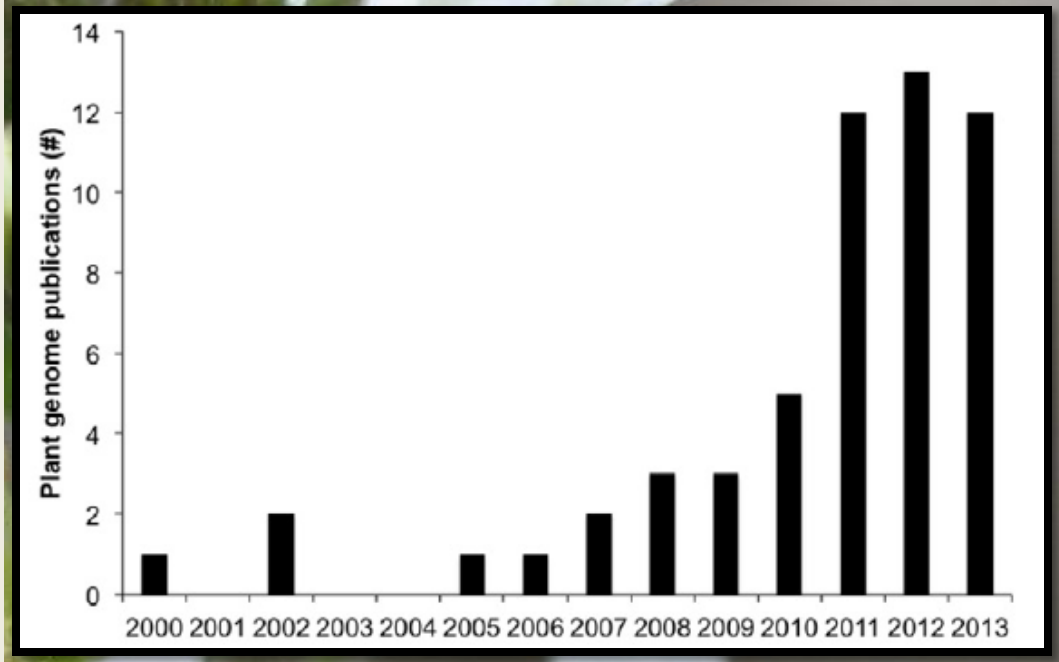


IT ALL STARTED WITH A WILD MUSTARD PLANT

(A) Number of plant genomes sequenced since *Arabidopsis thaliana* in 2000 by year.

❖Since the publication in 2000 of the model *Arabidopsis thaliana genome in the journal Nature, the number of* genomes has steadily increased, peaking in 2012 with 13 publications.

❖Genomes have been published in 12 different journals with 38 of the 55 (69%) published genomes appearing in Nature journals(*Nature, Nature Genetics, Nature Biotech, and Nature Communications); Science is second with six published* genomes.



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

Table 1. Published plant genomes.[†]

	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

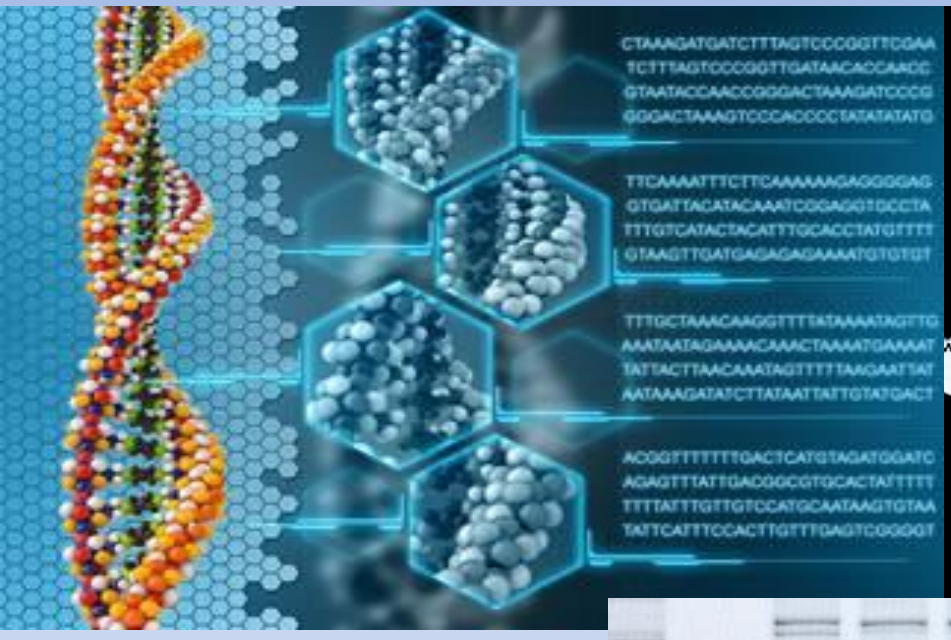
17	*Malus x domestica*	apple	2010	crop	dicot	17	742	604	81	57,386	67	1542	13	Sa,4	Nature Genetics	20802477
18	*Jatropha curcas*	jatropha	2010	crop	dicot	NA	380	286	75	40,929	37	NA	4	Sa,	DNA Research	21149391
19	*Theobroma cacao*	cocoa	2011	crop	dicot	10	430	327	76	28,798	24	473	20	Sa,4,I	Nature Genetics	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 cabbage	2011	crop	dicot	10	485	284	59	41,174	40	1971	27	I	Nature Genetics	21873998
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

Table 1. Continued.

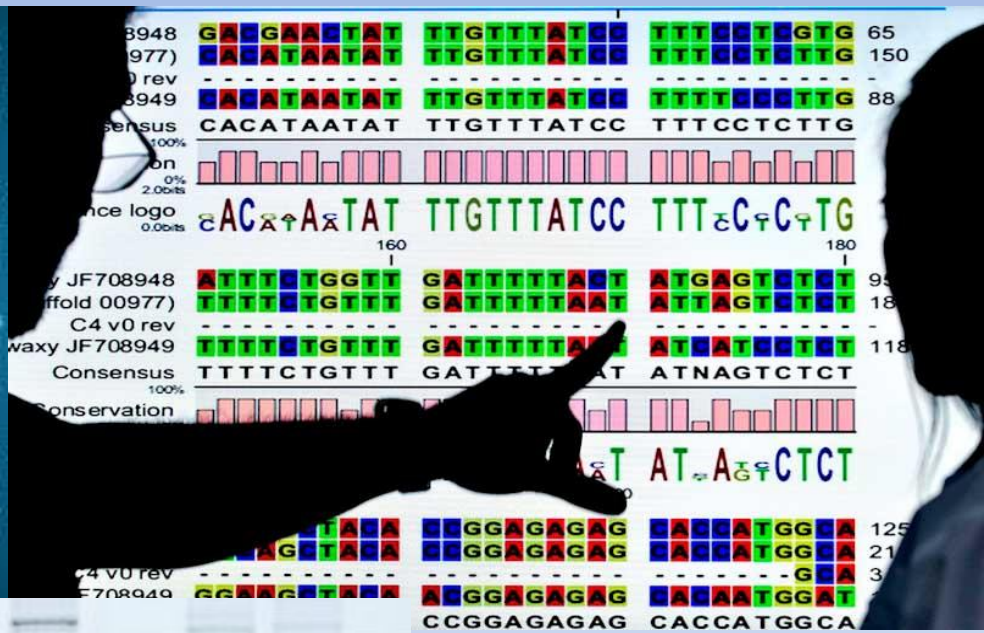
	Scientific name	Common name	Year	Type	Division or monocot/ dicot	Chr (#)	Size	Assembled	Assem	Gene (#)	Repeat	scaffold N50	contig N50	Sequencer types	Journal	PMID
							——Mb——		%		%	kb				
40	*Pyrus bretschneideri*	pear	2013	crop	dicot	17	527	512	97	42,812	53	541	36	I	Genome Research	23149293
41	*Citrullus 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
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

WHAT AFTER GENOME SEQUENCING?



GENOME SEQUENCE



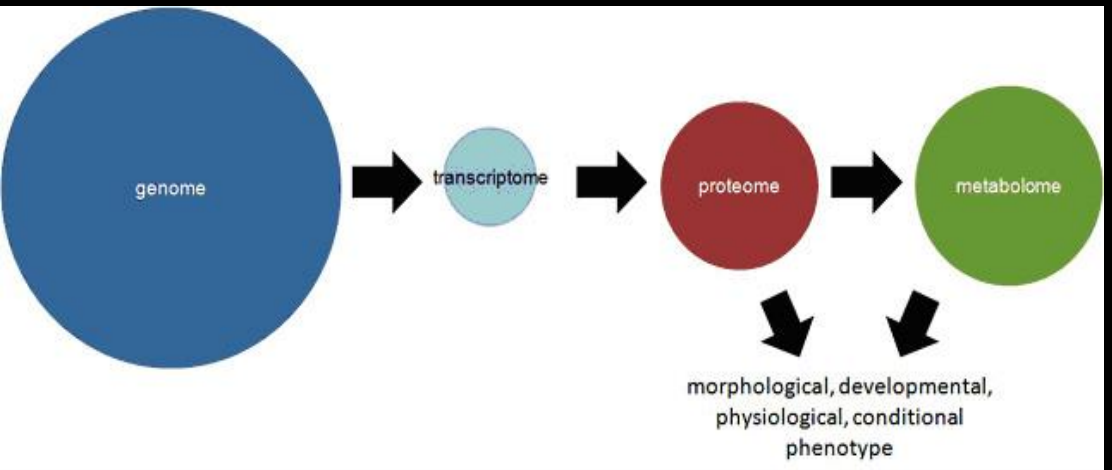
IDENTIFICATION OF GENES

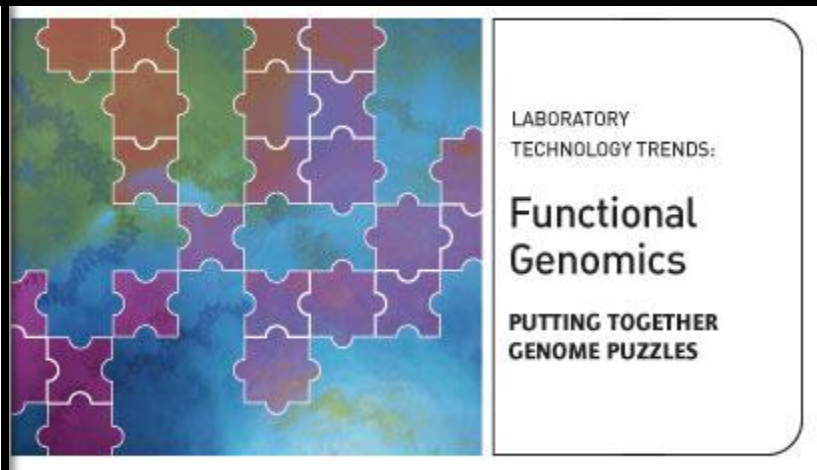
FUNCTIONAL ANNOTATION

FUNCTIONAL GENOMICS

Development and application of global (genome-wide or system wide) experimental approaches to assess gene functions by making use of the information and reagents provided by structural genomics.

Functional genomics includes a systematic analysis of mRNA and protein expression, exploration of gene product interactions and their influence on different phenotypic traits to define gene function. (Heiter and Beguski 1997)





FUNCTIONAL GENOMICS TECHNOLOGY

GENE LEVEL	TRANSCRIPTOME LEVEL	PROTEOME LEVEL
1. Homology searching	1. EST	1. 2D electrophoresis
2. Gene inactivation	2. SAGE	2. Differential display
3. Gene over-expression	3.DNA microarray	3. Yeast 2 hybrid system
4. Two step gene replacement	4. Subtractive hybridization	
5. Use of reporter gene expression		
6. Immuno-cytochemistry		

GENE LEVEL TECHNIQUES

- 1. HOMOLOGY SEARCHING

❖ Homologous genes share a common evolutionary ancestor, revealed by sequence similarity between genes.

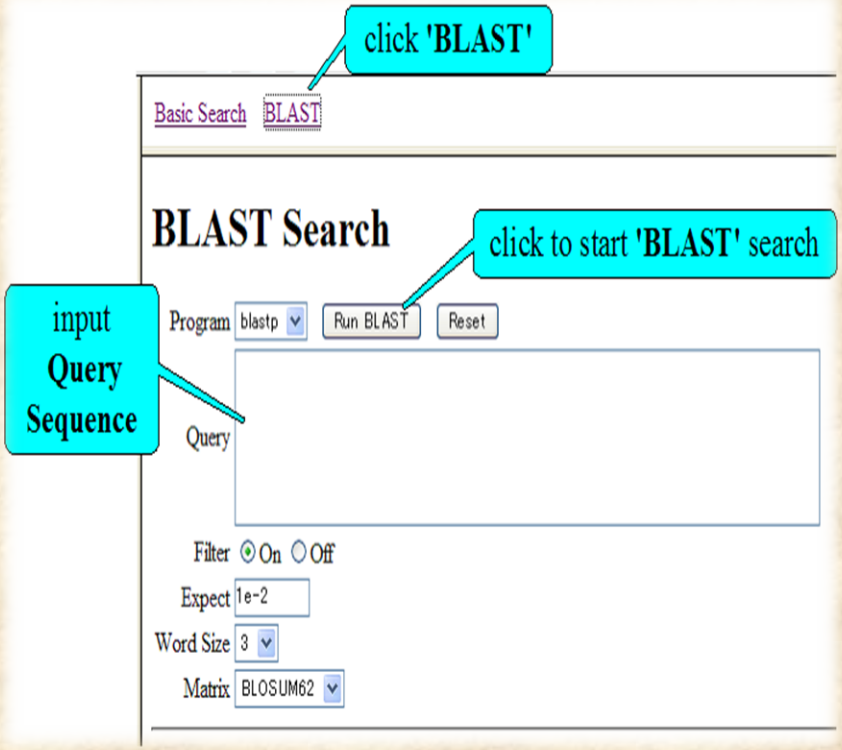
❖ A homology search can be conducted with a DNA sequence but usually it is first converted into an amino acid sequence before the search is carried out.

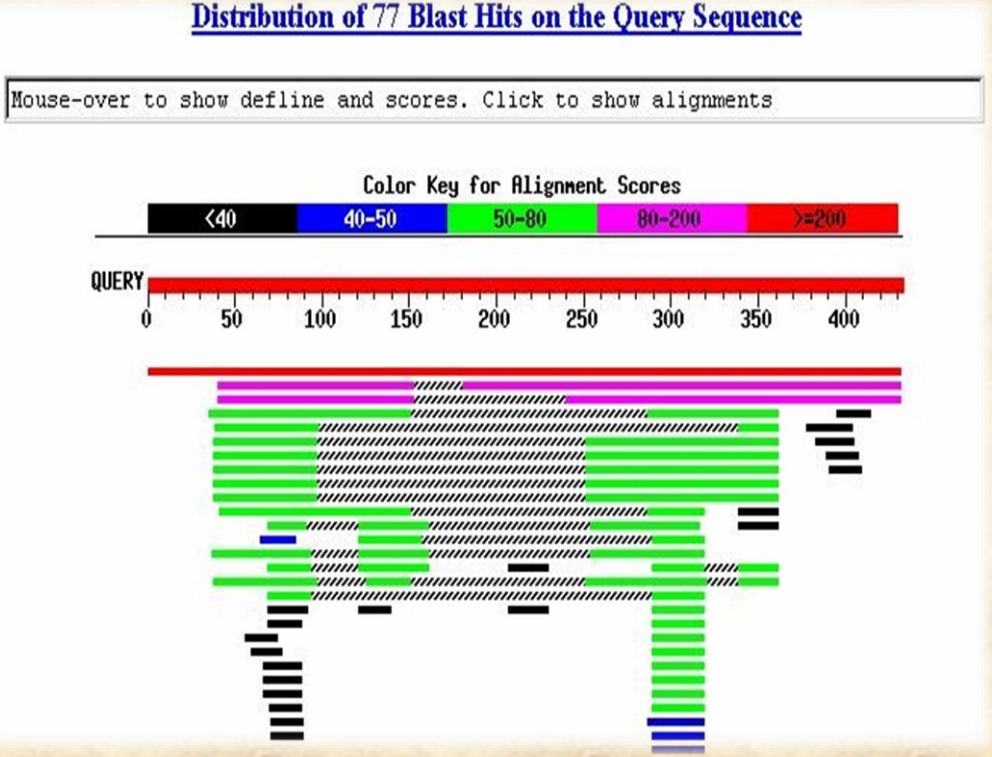
❖ Homology search is carried out using softwares such as **BLAST** (Basic Local Alignment Search Tool). The analysis can be carried out simply by logging on to the website for one of the DNA databases and entering the sequence into the online search tool.

❖ A modified version called **PSI-BLAST** (Position specific Iterated BLAST) can identify more distantly related sequences.

LIMITATIONS

1. A growing problem is the presence in the databases of genes whose stated functions are incorrect.
2. There are several cases where homologous genes have quite different biological functions.
3. There are also examples of genes that have similar sequences but no obvious evolutionary relatedness.

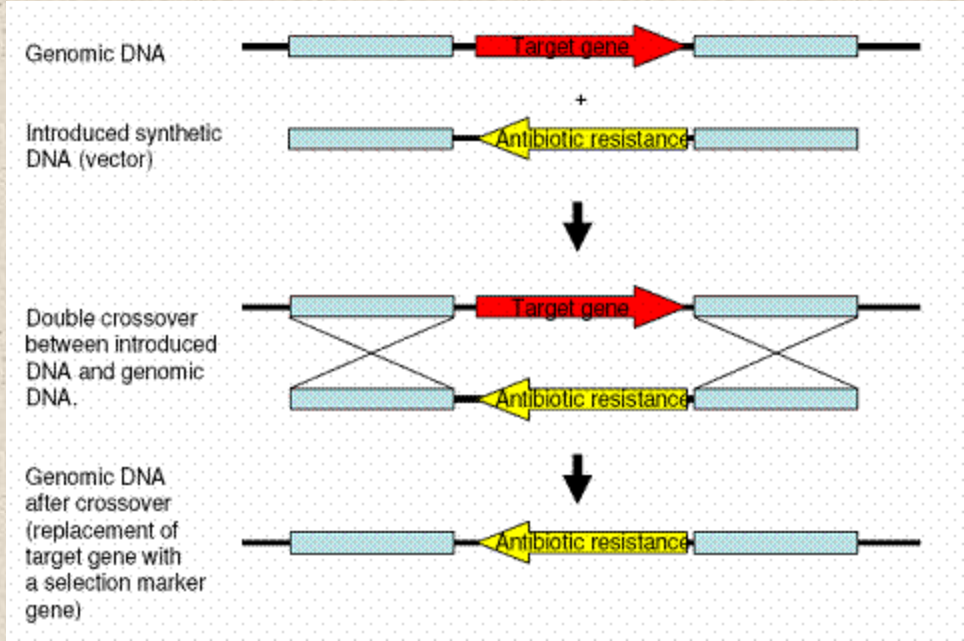




2. GENE INACTIVATION

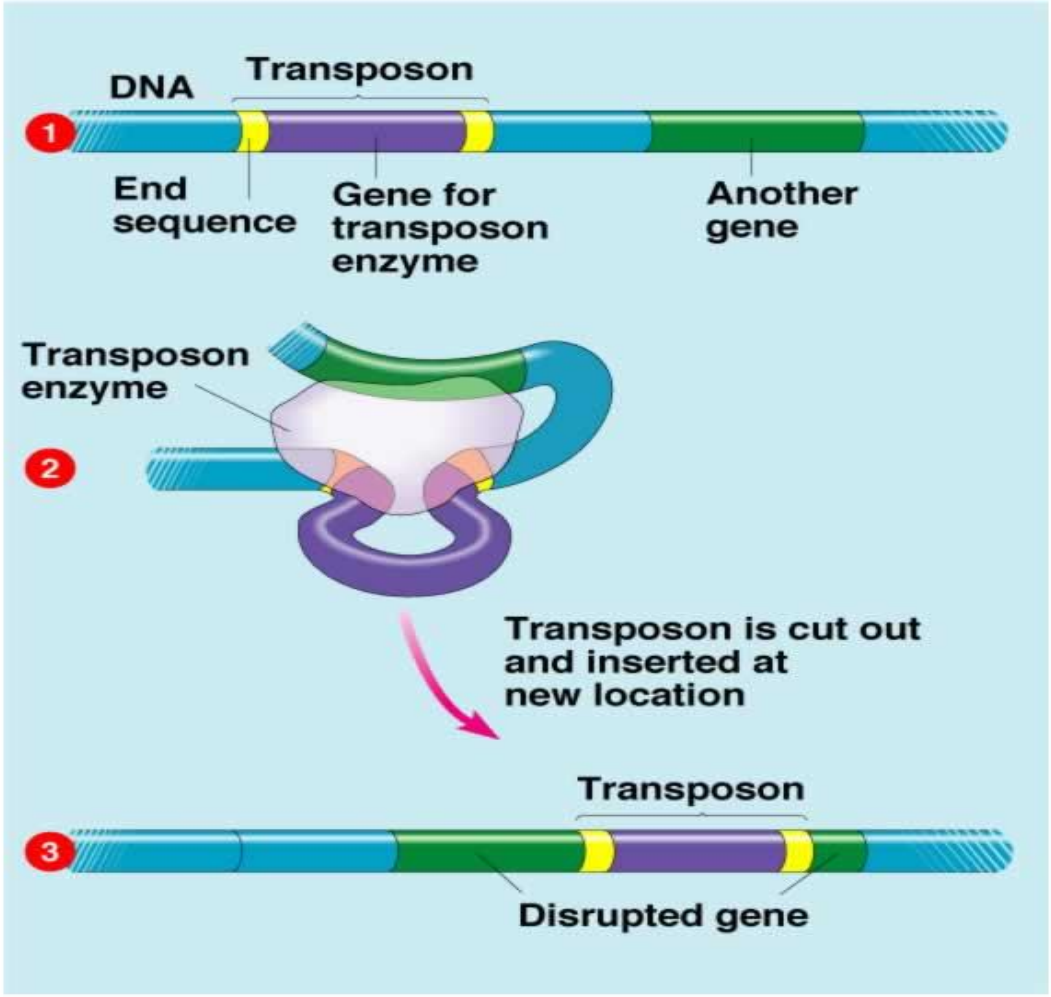
a) HOMOLOGOUS RECOMBINATION

The easiest way to inactivate a specific gene is to disrupt it with an unrelated segment of DNA. This can be achieved by homologous recombination between the chromosomal copy of the gene and a second piece of DNA that shares some sequence identity with the target gene.



b)TRANSPOSON TAGGING

- The inactivation is achieved by the insertion of a transposable element into the gene. Most genomes contain transposable elements and although the bulk of these are inactive, there are usually a few that retain their ability to transpose.
- Under normal circumstances, transposition is a relatively rare event but it is sometimes possible to use recombinant DNA techniques to make modified transposons that change their position in response to an external stimulus to get inserted into the target gene of interest thereby inactivating it.



c)RNA INTERFERENCE

- Long double stranded RNAs (dsRNAs typically >200nt) can be used to silence the expression of target genes in a variety of organisms and cell types (eg worms, fruit flies and plants).
- Upon introduction, the long dsRNA enter a cellular pathway that is commonly referred to as the RNA interference (RNAi) pathway.

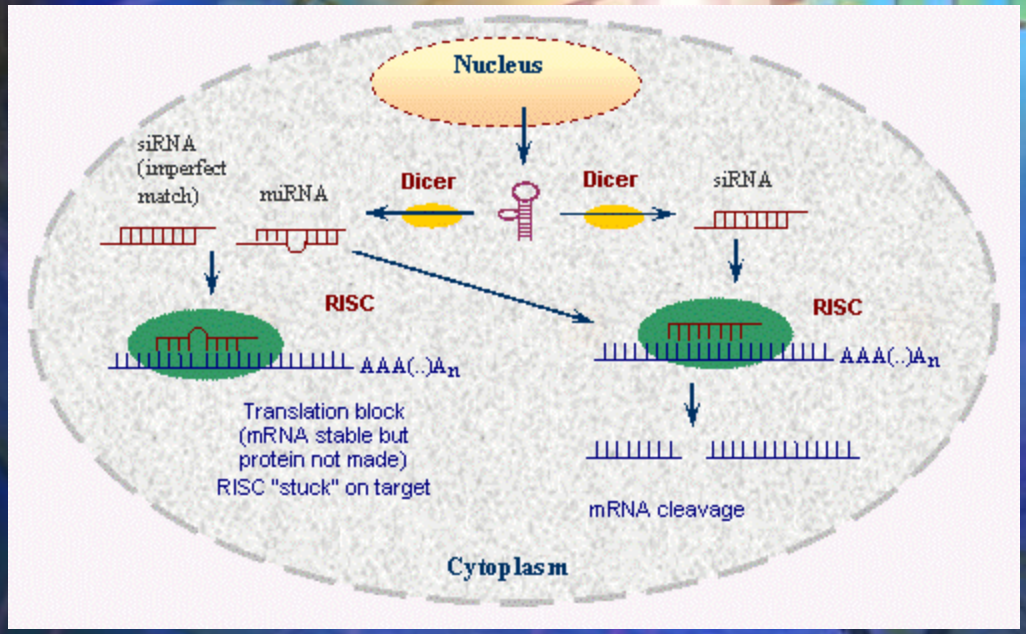
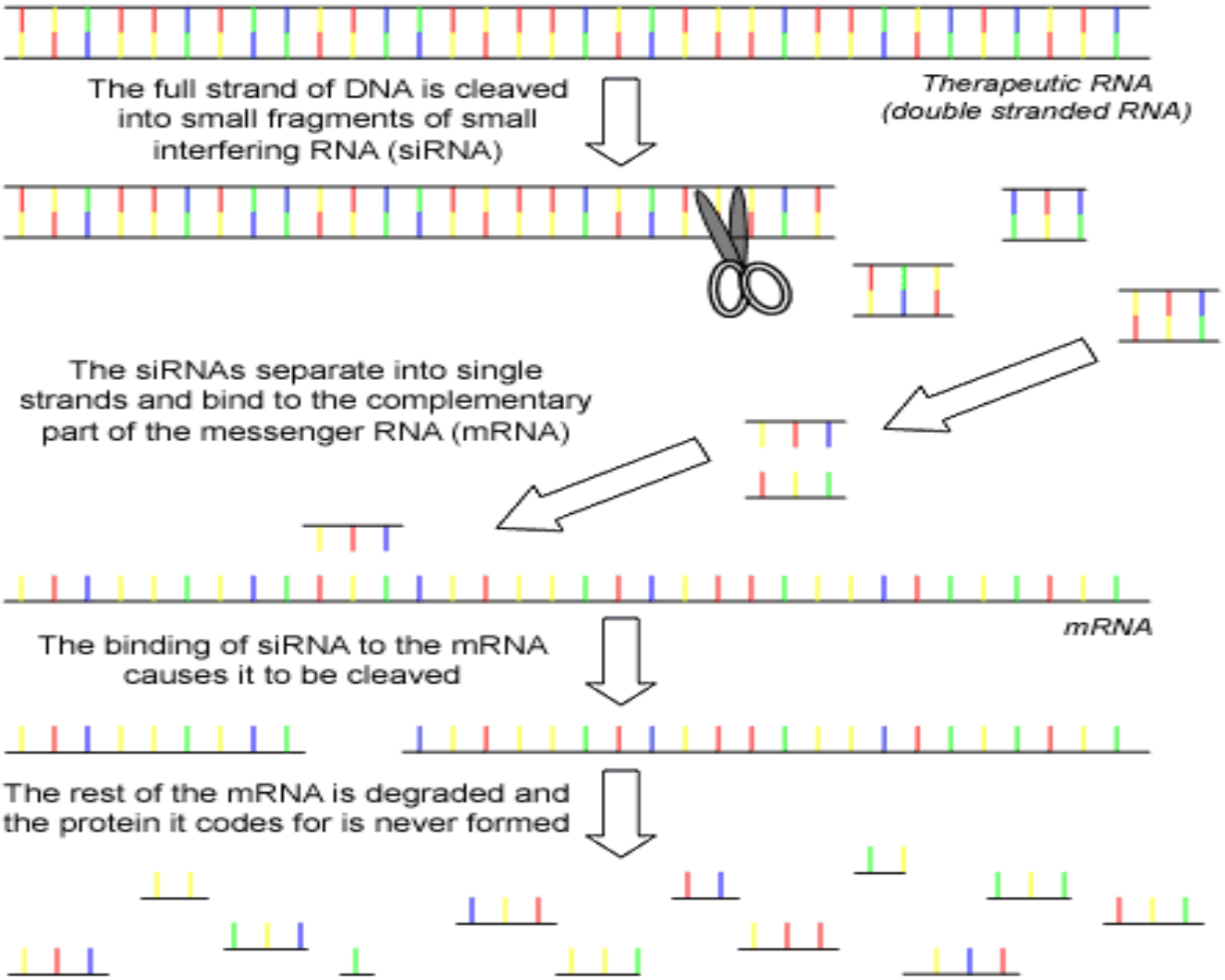


Figure H-2: RNA Interference - Simplified Mechanism



3. GENE OVER-EXPRESSION

- To over-express a gene a special type of cloning vector is used. The vector is designed to ensure that the cloned gene directs the synthesis of as much protein as possible.
- The vector is therefore multicopy meaning that it multiplies inside the host organism 40-200 copies per cell, so there are many copies of the test gene.
- The vector must also contain a highly active promoter so that each copy of the test gene is converted into large quantities of mRNA, again ensuring that as much protein as possible is made.

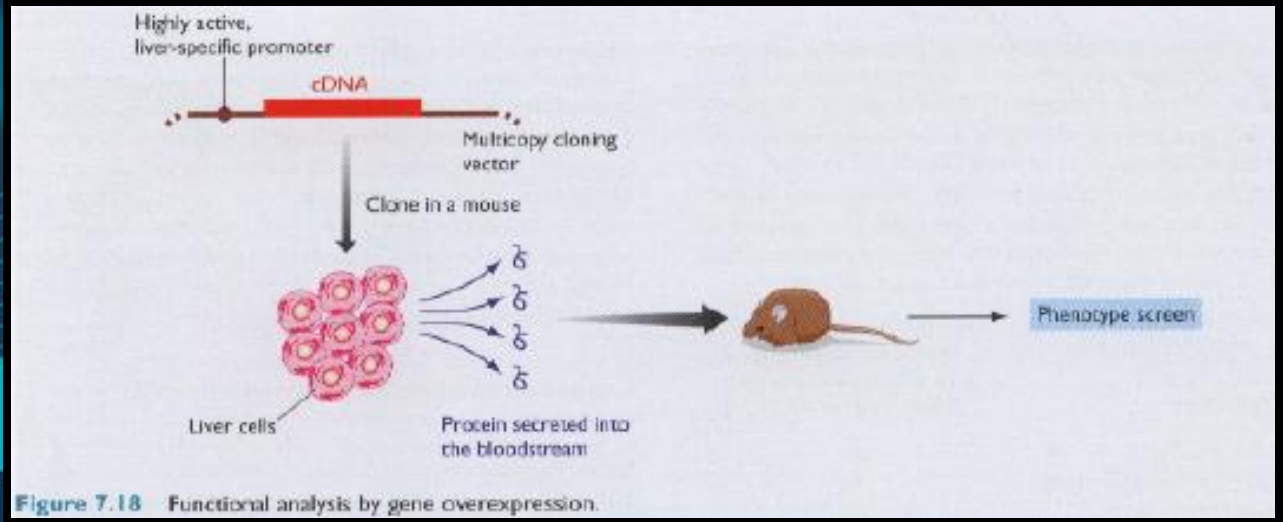


Figure 7.18 Functional analysis by gene overexpression.

4.TWO STEP GENE REPLACEMENT

- In this procedure the target gene is first replaced with the marker gene on its own, the cells in which this recombination takes place being identified by selecting for the marker gene phenotype.
- These cells are then used in the second stage of the gene replacement, when the marker gene is replaced by the mutated gene, success now being monitored by looking for cells that have lost the marker gene phenotype.
- These cells contain the mutated gene and their phenotypes can be examined to determine the effect of the directed mutation on the activity of the protein product.

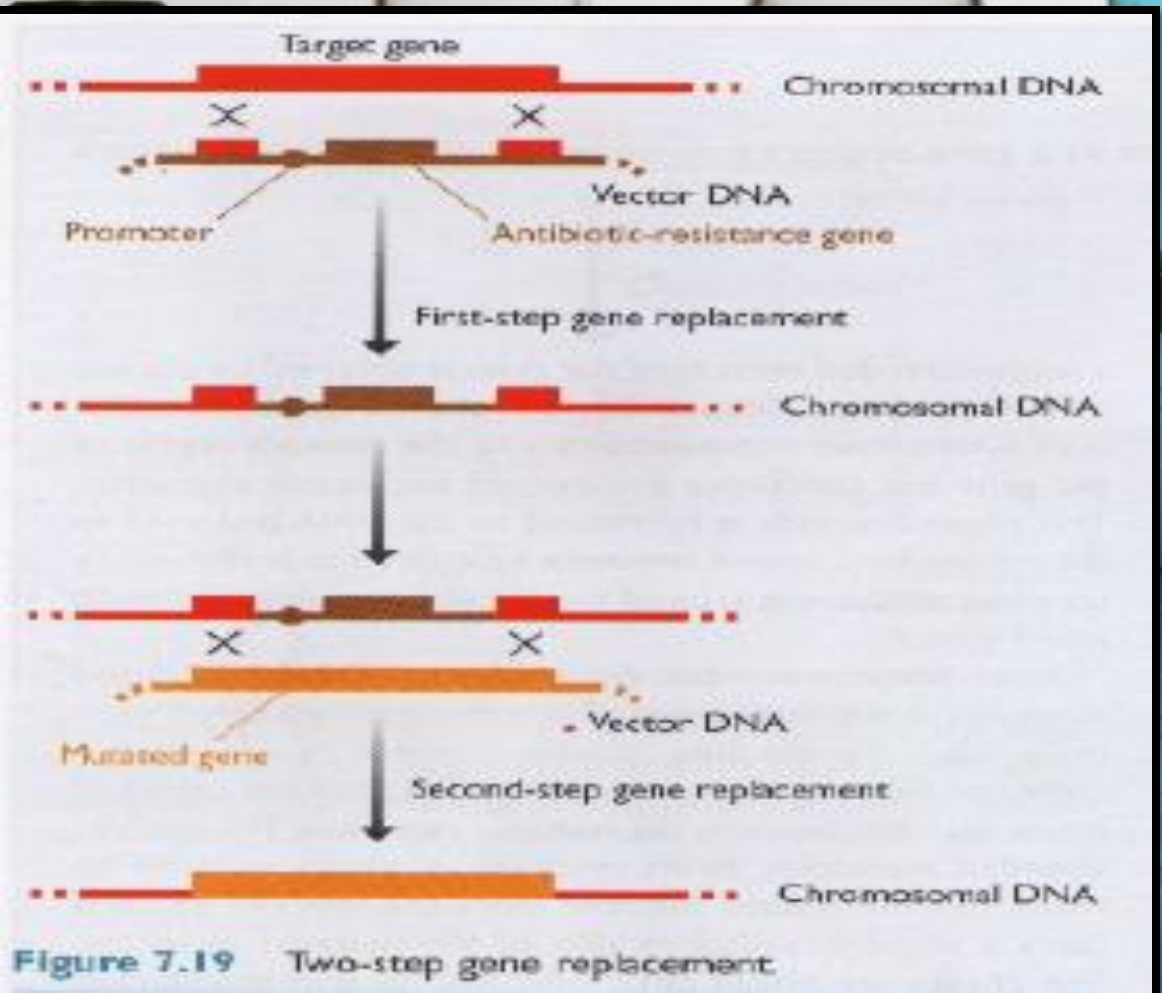


Figure 7.19 Two-step gene replacement

5.USE OF REPORTER GENE

- For the reporter gene to give a reliable indication of where and when a test gene is expressed, the reporter must be subjected to the same regulatory signals as the test gene. This is achieved by replacing the ORF of the test gene with the ORF of the reporter gene.
- Most of the regulatory signals that control gene expression are contained in the region of DNA upstream of the ORF, so the reporter gene should now display the same expression pattern as the target gene.
- The expression pattern can therefore be determined by examining the organism for the reporter signal.

Table 7.1 Examples of reporter genes

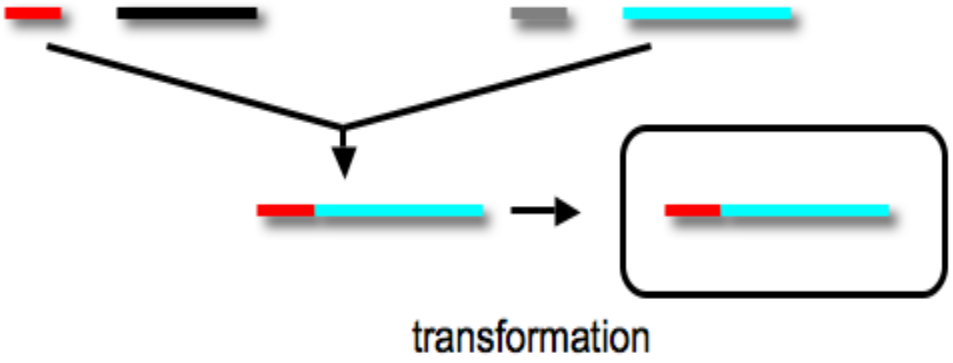
Gene	Gene product	Assay
lacZ	β-galactosidase	Histochemical test
uidA	β-glucuronidase	Histochemical test
lux	Luciferase	Bioluminescence
GFP	Green fluorescent protein	Fluorescence

Inducible promoter

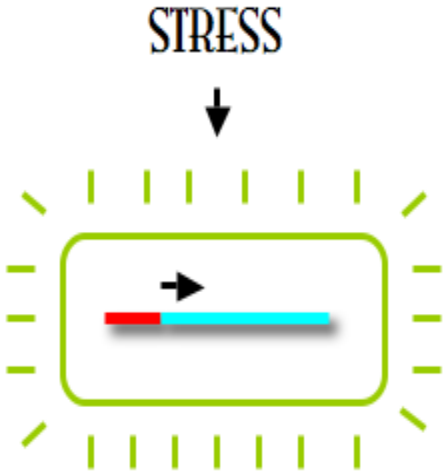
- Resistance gene
- SOS gene
- Shock response gene
- etc.

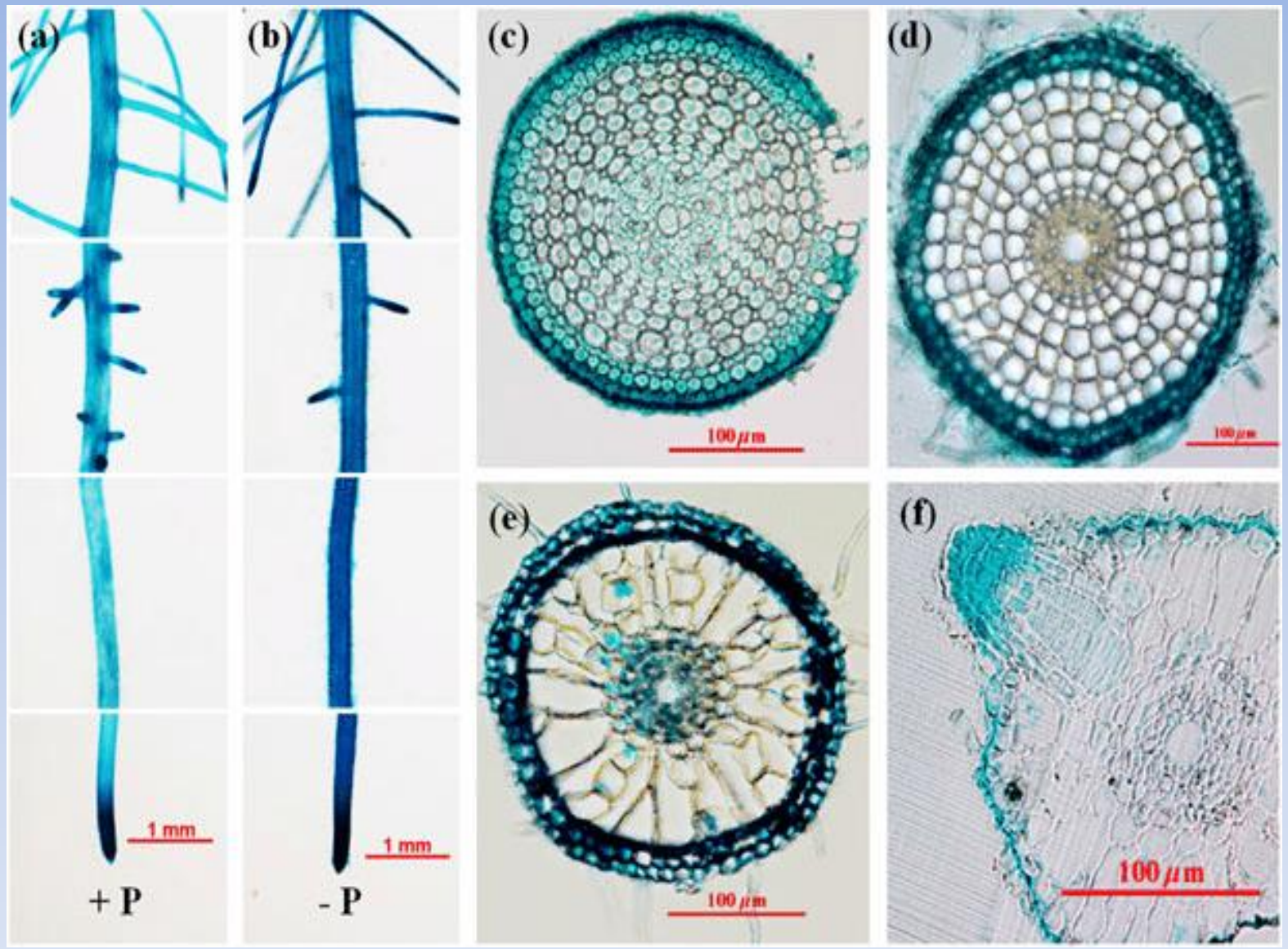
Reporter

- Luciferase (*lux*, *luc*)
- Green fluorescent protein
- ß-galactosidase
- ß-glucuronidase
- etc.



transformation





GUS reporter gene expression in transgenic plants

6. IMMUNO-CYTOCHEMISTRY

- The only way to determine where the protein is located is to search for it directly. This is done by immuno-cytochemisty which makes use of antibody that is specific for the protein of interest and so binds to this protein and no other.
- The antibody is labeled so that its position in the cell, and the position of the target protein, can be visualized.

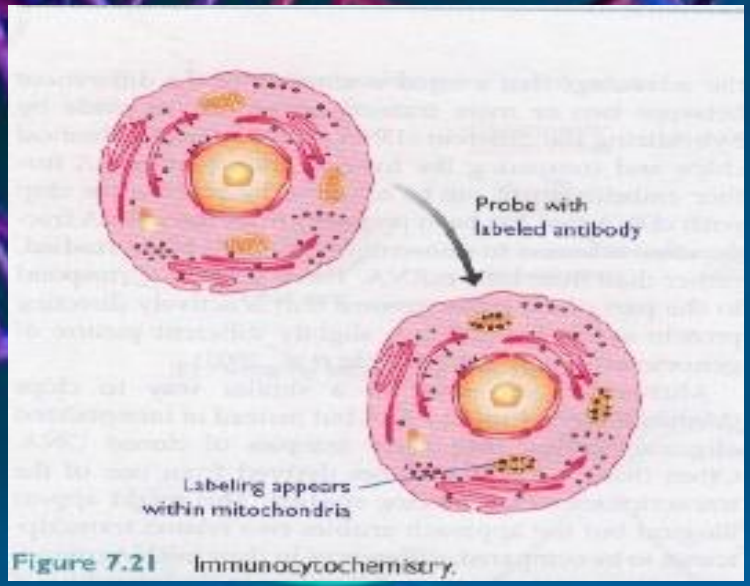


Figure 7.21 Immunocytochemistry.

TRANSCRIPTOME LEVEL

1.EXPRESSED SEQUENCE TAGS

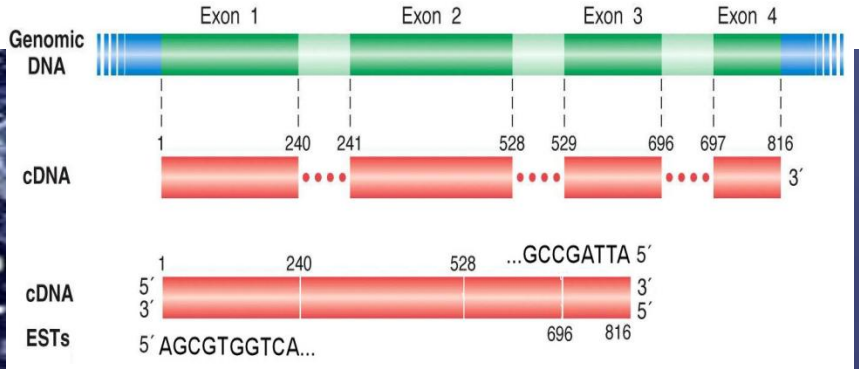
ESTs are short (200-500 nucleotides) DNA sequences that can be used to identify a gene that is being expressed in a cell at a particular time.

In this procedure we isolate the messenger RNA (mRNA) from a particular tissue.

It is treated with reverse transcriptase. This produces a complementary DNA. The cDNA differs from the normal gene in lacking the intron sequences.

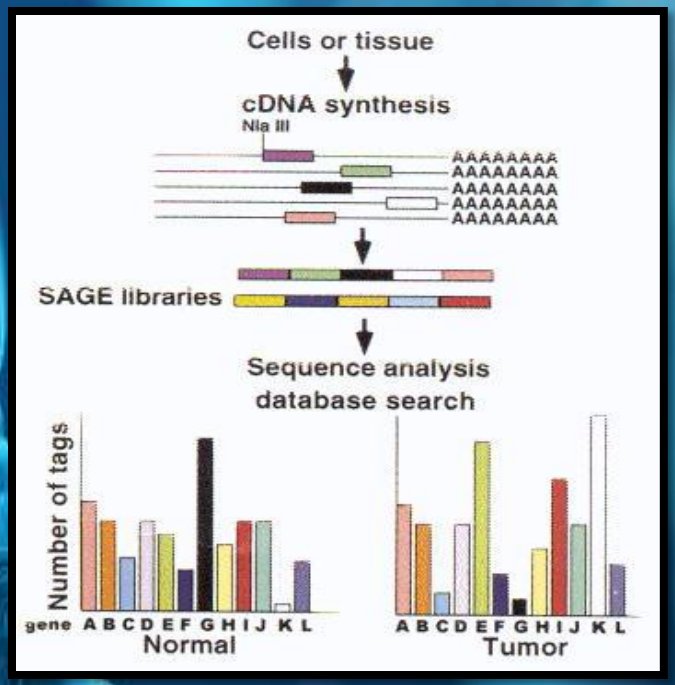
Sequence 200-500 nucleotides at both the 5’ and 3’ ends of each cDNA.

Examine the database of the organism’s genome to find a matching sequence. That is the gene that was expressed.



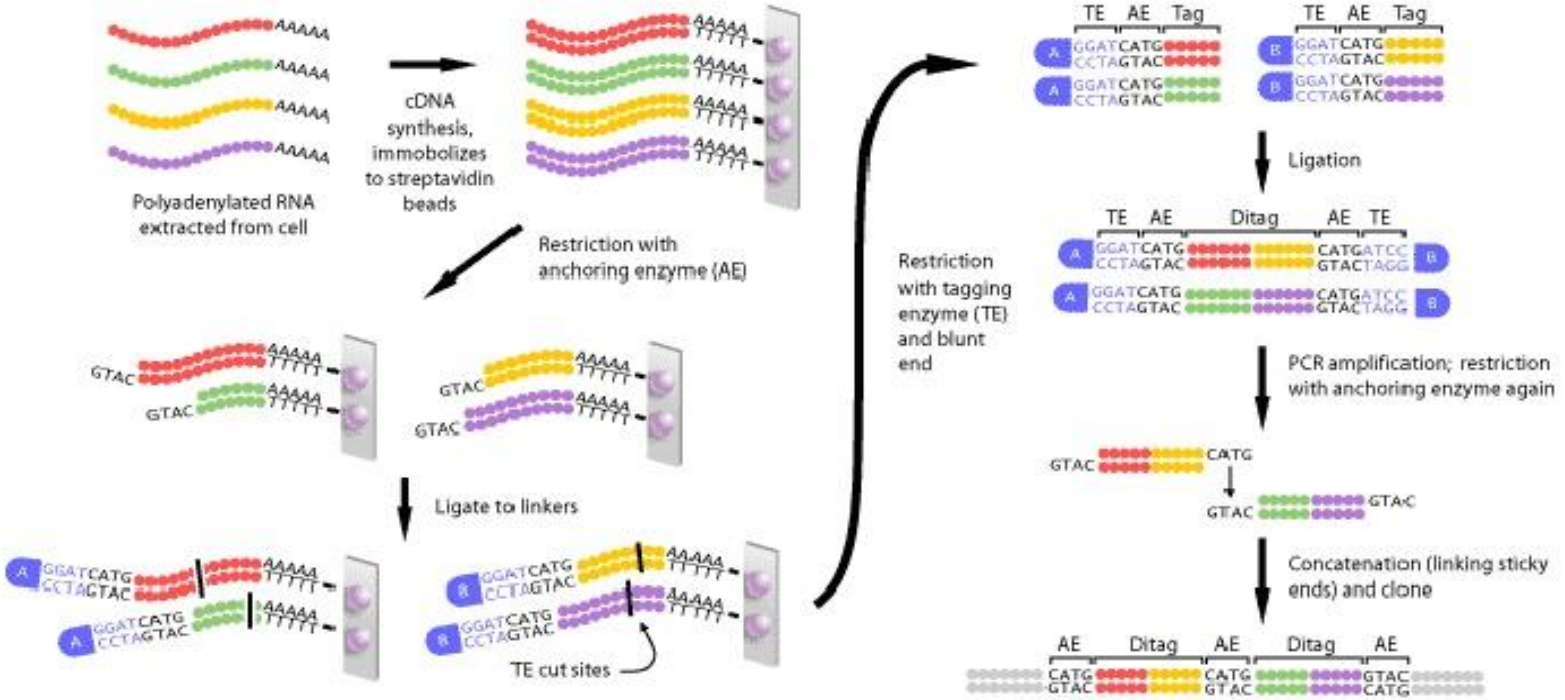
2. SAGE

- Serial analysis of gene expression (SAGE) involves digesting cDNA to generate short (~12 base pair), gene-specific sequence tags that represent the transcriptome of a cell population, followed by sequencing and quantification of these tags.
- SAGE does not require prior knowledge of target sequences and provides a quantitative gene expression.



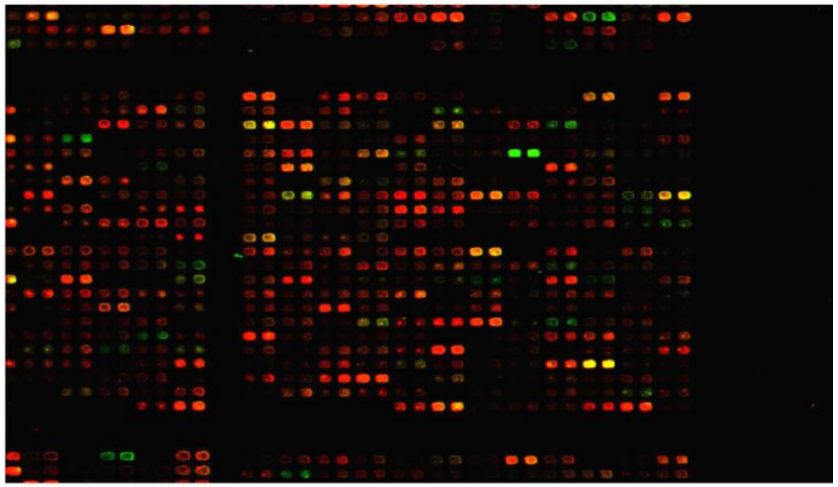
SAGE
Serial Analysis of Gene Expression

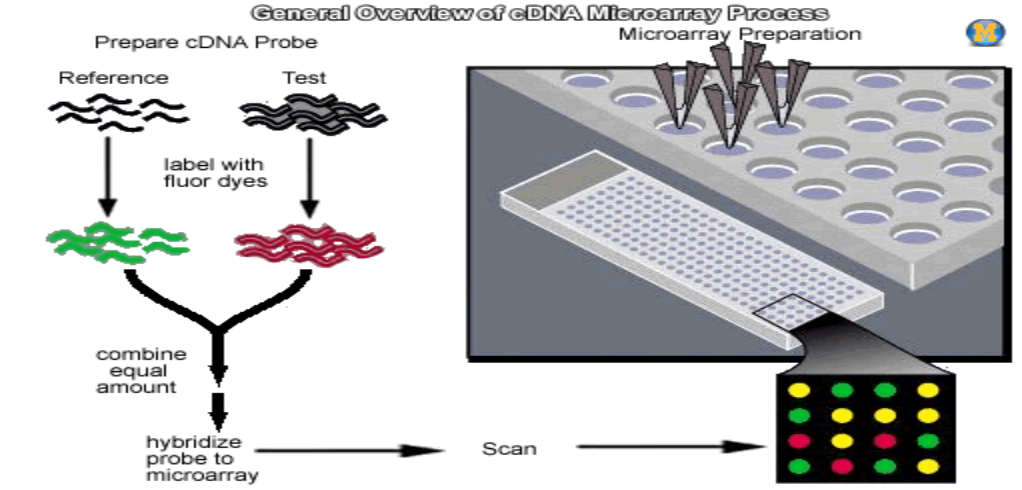
Serial analysis of gene expression (SAGE)



3. DNA MICROARRAY

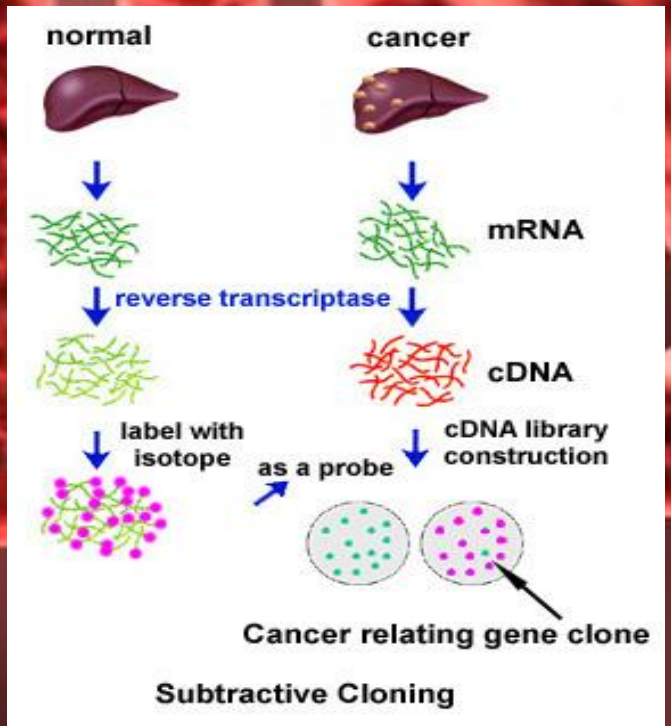
- A microarray is a set of short Expressed Sequence Tags made from a cDNA library of a set of known (or partially known) gene loci.
- The ESTs are spotted onto a cover-slip-sized glass plate.
- A cell's RNA is extracted. This RNA is then multiplied, labeled with fluorescence and hybridized to existing DNA on the microarray. After hybridization, the probes that were hybridized with targets are fluorescent and a computer scanner is able to detect this fluorescence.
- Those probes that are fluorescent correspond to the genes that were expressed in the cell.





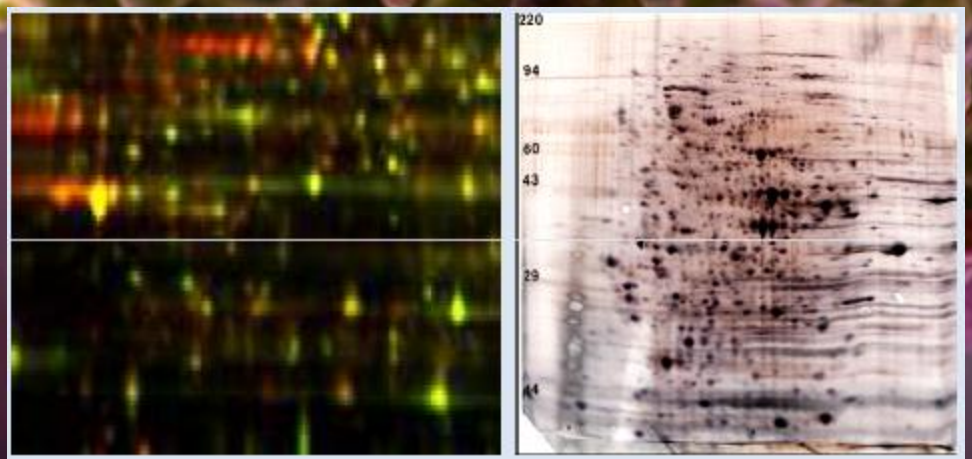
4. SUBTRACTIVE HYBRIDIZATION

- Subtractive hybridization is used to characterize transcripts that are differentially expressed between two cell populations.
- The cDNA synthesized from one population is mixed with mRNA form another population, which results in the formation of cDNA-mRNA hybrids of transcripts that are expressed in both the populations.
- More abundant mRNA remains unhybridized, and can then be isolated and characterized.

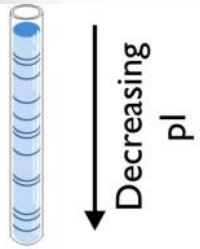


PROTEOME LEVEL

- 1. 2D GEL ELECTROPHORESIS
- PAGE is the standard method for separating proteins, but the usual procedure, in which proteins are separated according to their molecular weights, is unable to resolve the many proteins in an average proteome.
- To separate individual proteins, the polyacrylamide gel is rotated by 90˚and a second electrophoresis is carried out at right angles to the first.



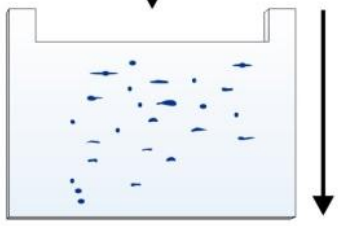
1. Separation of proteins by pI value



Decreasing pI

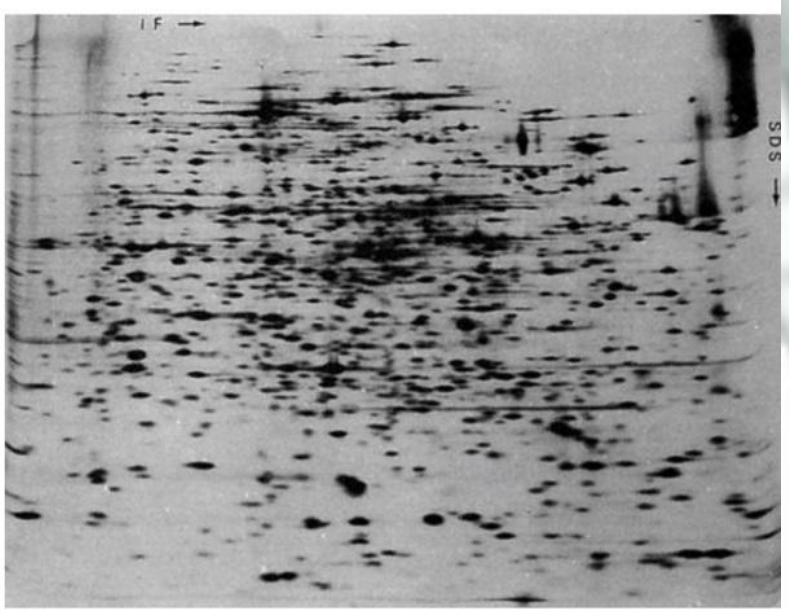
2. Soaking the gel in SDS solution and fitting it on an SDS PA gel

3. Separating the proteins by molecular mass with SDS PAGE



Decreasing molecular mass

Decreasing pI



Approximately 1000 *E. coli* proteins on a single 2D gel

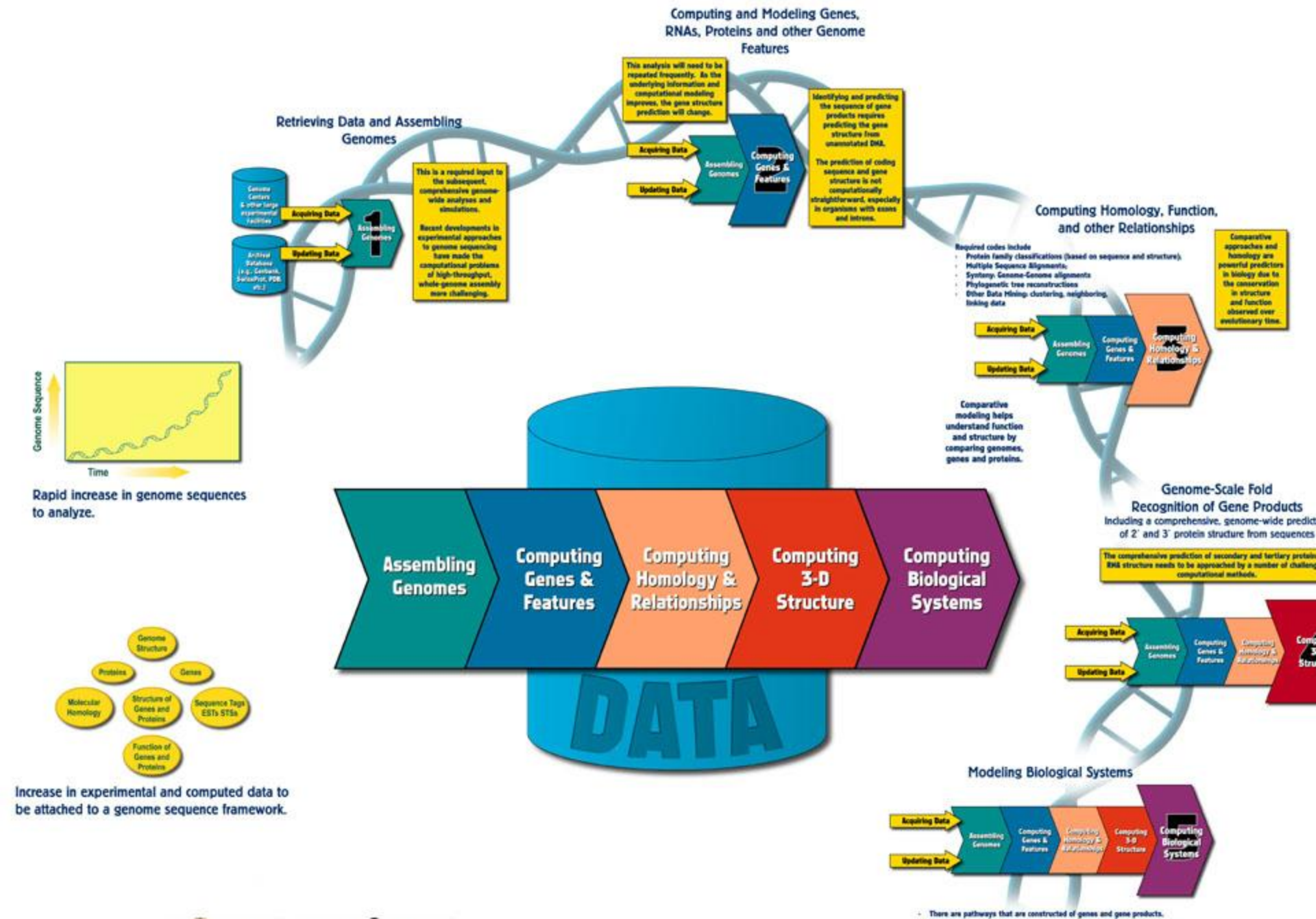
2. DIFFERENTIAL DISPLAY

- In phage display a special type of cloning vector is used, one based on λ bacteriophage or one of the filamentous bacteriophages such as M13.
- The vector is designed so that a new gene that is cloned into it is expressed in such a way that its protein product becomes fused with one of the phage coat proteins.
- The phage protein therefore carries the foreign protein into the phage coat, where it is displayed in a form that enables it to interact with other proteins that the phage encounters.

3. YEAST TWO HYBRID SYSTEM

- The two-hybrid system is a molecular genetic tool which facilitates the study of protein-protein interactions. If two proteins interact, then the reporter gene(gal1-lacZ) is transcriptionally activated.
- The intact Gal4 protein is a transcriptional activator which has two separate functions (a DNA-binding and an activating domain). This application of knowledge is achieved by splitting the yeast Gal4 gene in two parts- the activating domain and the binding domain.
- The split protein does not work unless the two parts are in physical contact.

Computing the Genome Revolution: Biology for the 21st Century



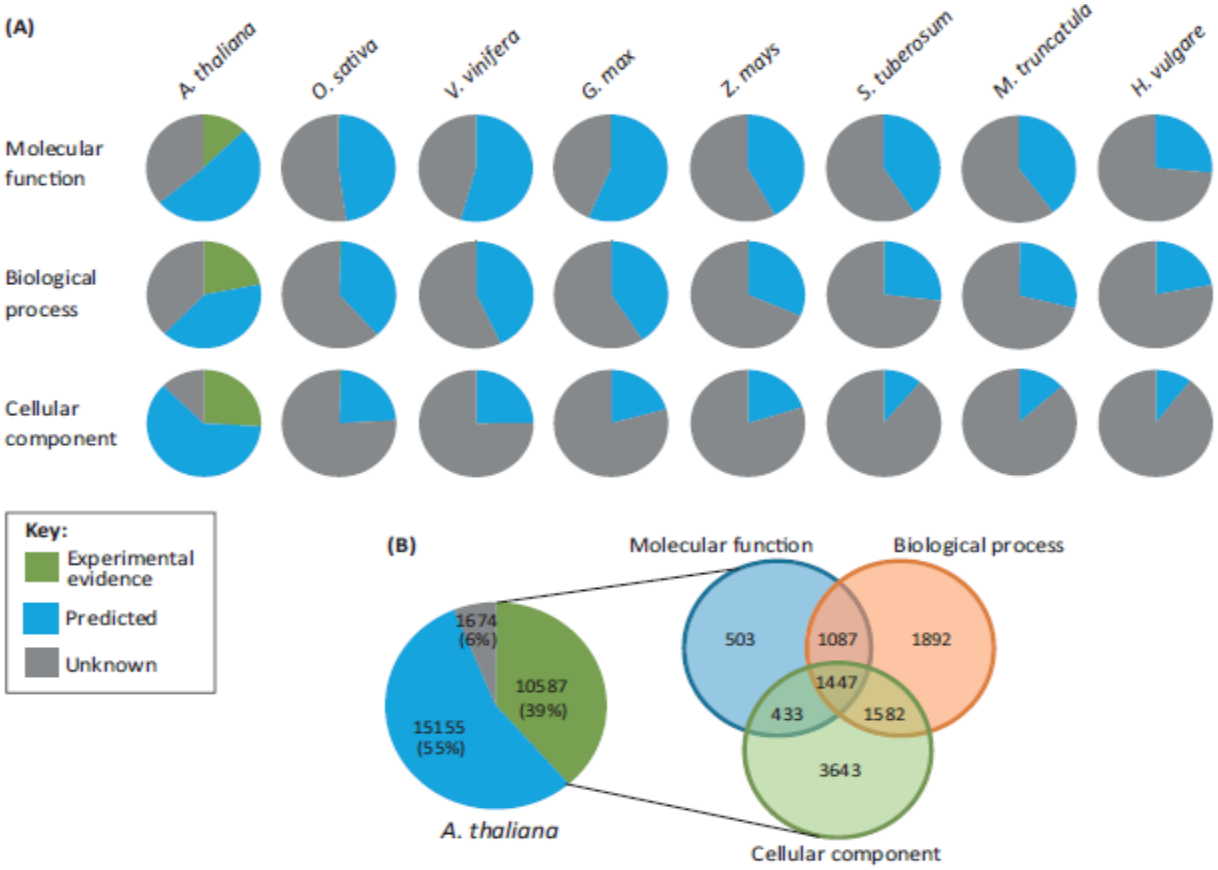
CellPress

Special Issue: Systems Biology

Towards revealing the functions of all genes in plants

Seung Yon Rhee[1] and Marek Mutwil[2]

[1] Carnegie Institution for Science, Department of Plant Biology, 260 Panama St, Stanford, CA 94305, USA
[2] Max Planck Institute for Molecular Plant Physiology, 14476 Potsdam, Germany



How little we know

The elucidation of the genome sequence of many organisms, one of the outstanding achievements of our generation, confirmed what most biologists already suspected – that we know little about what most genes do. For example, approximately 40% of *Arabidopsis* (*Arabidopsis thaliana*, thale cress) and 1% of rice (*Oryza sativa*) protein-coding genes have had some aspect of their functions annotated based on experimental evidence (Figure 1) [1,2]. Moreover, we know about the biochemical activity, subcellular location, and biological role of only ~5% of *Arabidopsis* genes based on experimental evidence. It is difficult to determine the number of experimentally characterized genes in public databases for any plant species other than for *Arabidopsis* and rice. This paucity and disparity in the level of functional annotation in different plant species is a bottleneck for understanding how biological processes are organized, how they function, and how they evolved in plants.

Table 1. Availability and source of omics data relevant for predicting gene function

Plant species	Genomes (year published)[a]	Transcriptome RNA samples[b]		Protein interactions[c]	Genetic interactions[c]	3D structures[d]
		Microarray	RNAseq			
Arabidopsis thaliana	2000	26 747	391	16 697	171	2135
Oryza sativa	2002	5464	85	0	0	129
Vitis vinifera	2007	1064	7	0	0	19
Hordeum vulgare	2009	2558	4	0	0	111
Medicago truncatula	2009	1184	0	0	0	15
Zea mays	2009	3275	58	1	0	182
Glycine max	2010	1565	67	0	0	136
Solanum tuberosum	2011	1061	16	0	0	32

[a]Data from Genomes OnLine database (http://www.genomesonline.org/).

[b]Data from ArrayExpress (http://www.ebi.ac.uk/arrayexpress/): downloaded June 16, 2013.

[c]Data from BioGRID (http://thebiogrid.org/).

[d]Data from NCBI (http://www.ncbi.nlm.nih.gov/structure/).

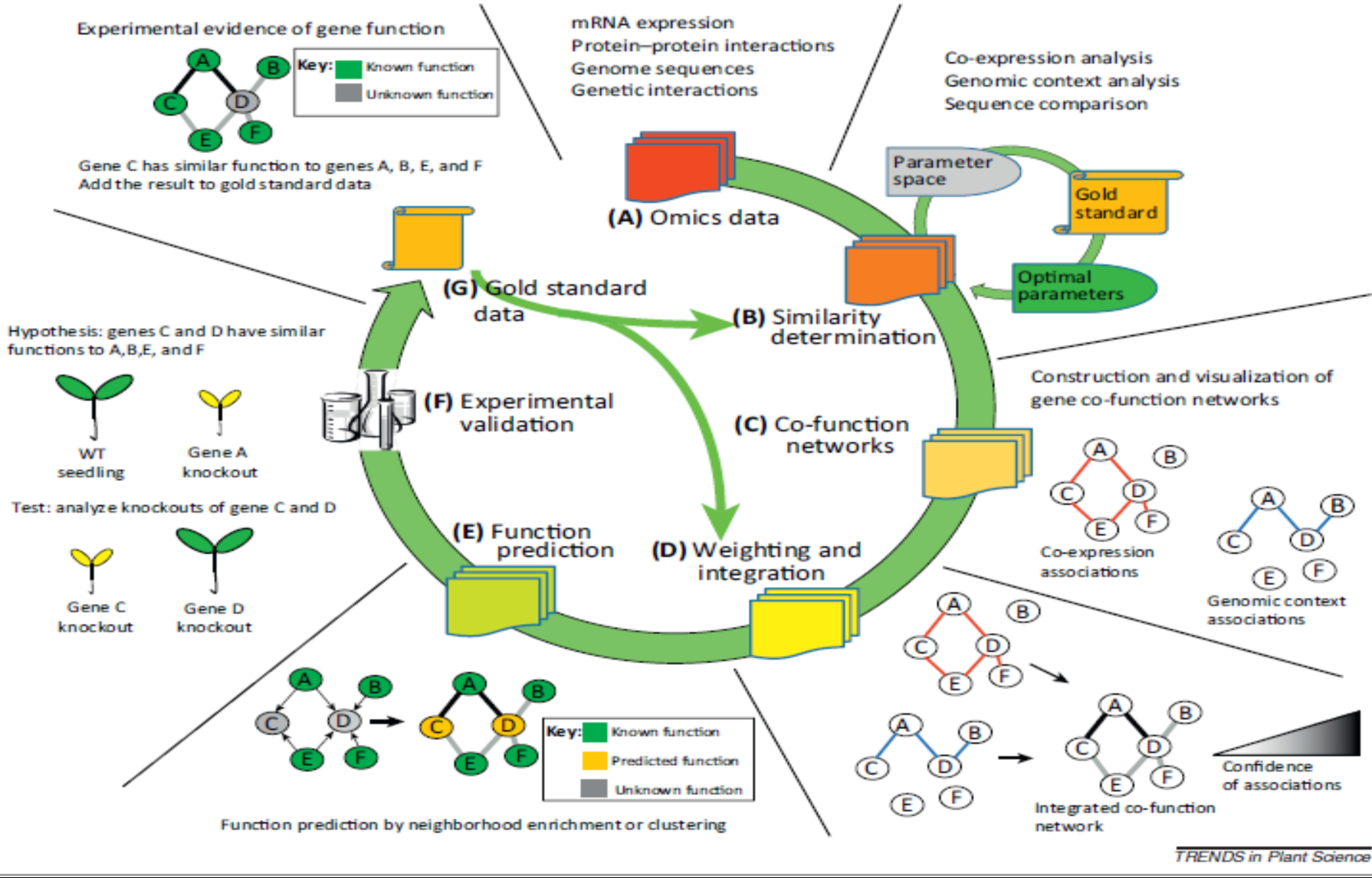


Figure 2. Process of systematic function elucidation using omics data. **(A)** Omics datasets are generated. **(B)** The data are analyzed by appropriate methods to determine similarities between genes. Gold standard data are used to find optimal parameter values and to train machine learning algorithms. **(C)** Each method generates a matrix of gene–gene associations and can be visualized as a network. **(D)** To integrate results from the different methods, gold standard data are used to weight the gene–gene associations. Associations consistent with gold standard data are assigned more weight, whereas poorly performing associations are assigned less weight. Weighted associations from the different data sources are integrated into one co-function network. **(E)** The network can now be used to predict gene function by using neighborhood enrichment, clustering, or other methods. **(F)** Predictions are used for focused reverse genetic testing, where uncharacterized genes can be associated with a biological process of interest. **(G)** New evidence from experimental validation is used to expand the gold standard data, which can be used to train and improve future predictions.

Table 2. Tools useful for inferring the biological processes of plant genes

Tool	Network visualization	New member identification	Function prediction	Data types use[m]	Data integration scheme	Confidence of association or prediction shown[m]	Tool performance published[m]	Refs
AraNet[a]	Y	Y	Y	SS, CE, PPI, GI, GC	Bayesian	Y	Y	[12]
AFMSD[b]	N	Y	Y	SS, CE, PPI, GI, GC	Clustering	Y	Y	[92]
ATTED-II[c]	Y	Y	N	SS, CE	None	Y	Y	[93]
BMRF[d]	N	Y	Y	SS, CE, PPI	Bayesian	N	Y	[94]
CoP[e]	Y	Y	N	SS, CE	None	Y	N	[95]
GeneMANIA[f]	Y	Y	Y	SS, CE, PPI, GI	Support vector machine	Y	Y	[55]
GO-At[g]	N	Y	Y	SS, CE, PPI, GC	Bayesian	N	Y	[96]
PlaNet[h]	Y	Y	N	SS, CE	None	Y	Y	[97]
SCoPNet[i]	Y	Y	Y	CE	None	N	N	[98]
StarNet2[j]	Y	Y	N	SS, CE	None	Y	N	[99]
STRING[k]	Y	Y	N	SS, CE, PPI, GI, GC	None	Y	N	[100]
VirtualPlant[l]	Y	Y	N	SS, CE	None	N	N	[101]

[a]http://www.functionalnet.org/aranet/.

[b]http://bioinformatics.psb.ugent.be/cig_data/plant_modules/.

[c]http://atted.jp/.

[d]http://www.ab.wur.nl/bmrf/.

[e]http://webs2.kazusa.or.jp/kagiana/cop0911/.

[f]http://www.genemania.org/.

[g]http://www.bioinformatics.leeds.ac.uk/goat.

[h]http://aranet.mpimp-golm.mpg.de/.

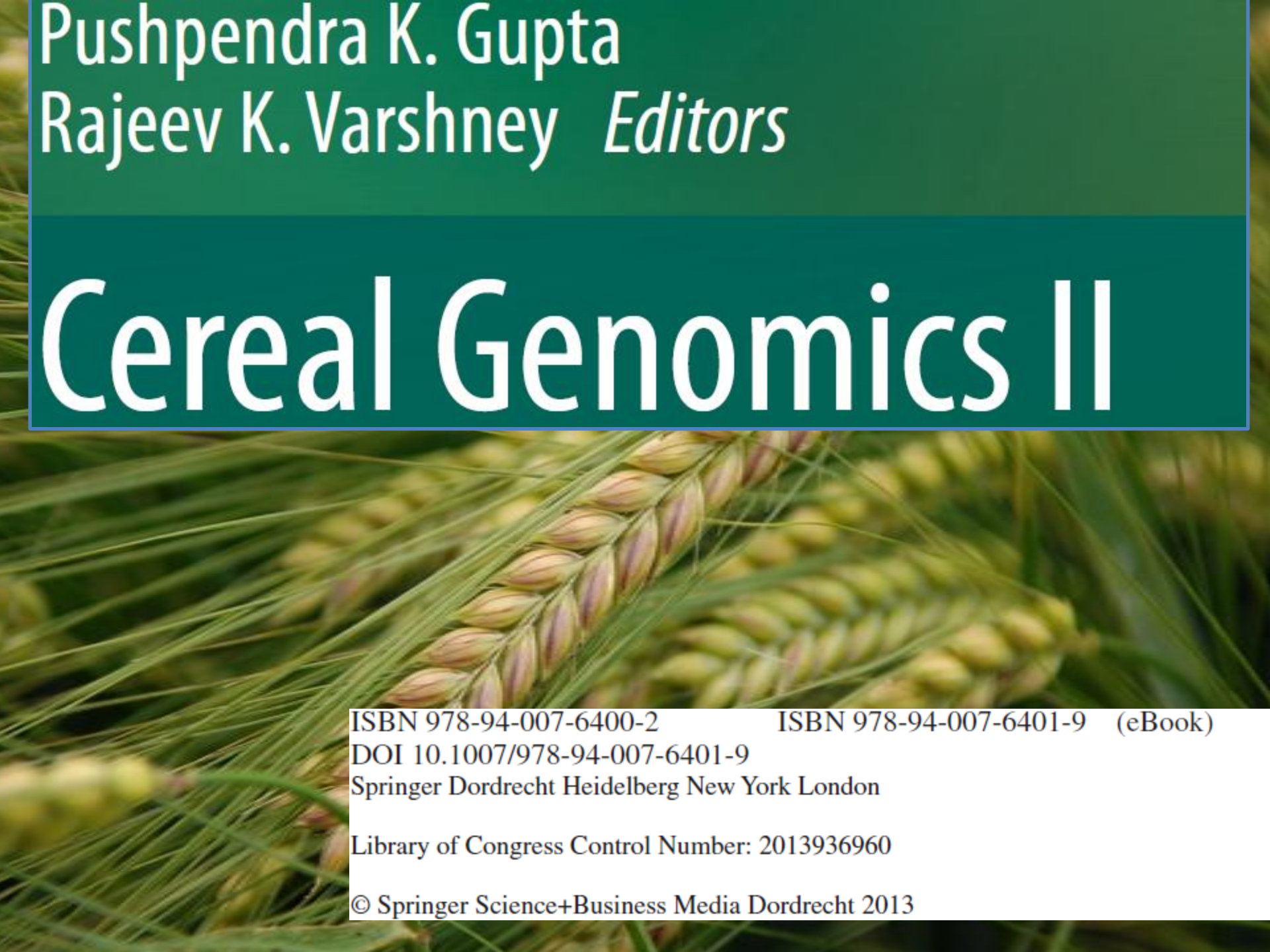
[i]http://bree.cs.nott.ac.uk/arabidopsis/neighbor/network.php.

[j]http://vanburenlab.medicine.tamhsc.edu/starnet2.html.

[k]http://string-db.org/.

[l]http://www.virtualplant.org/.

[m]Abbreviations: CE, co-expression; GC, genomic context; GI, genetic interaction; N, no; PPI, protein–protein interaction; SS, sequence similarity; Y, yes.



ISBN 978-94-007-6400-2 ISBN 978-94-007-6401-9 (eBook)
DOI 10.1007/978-94-007-6401-9
Springer Dordrecht Heidelberg New York London

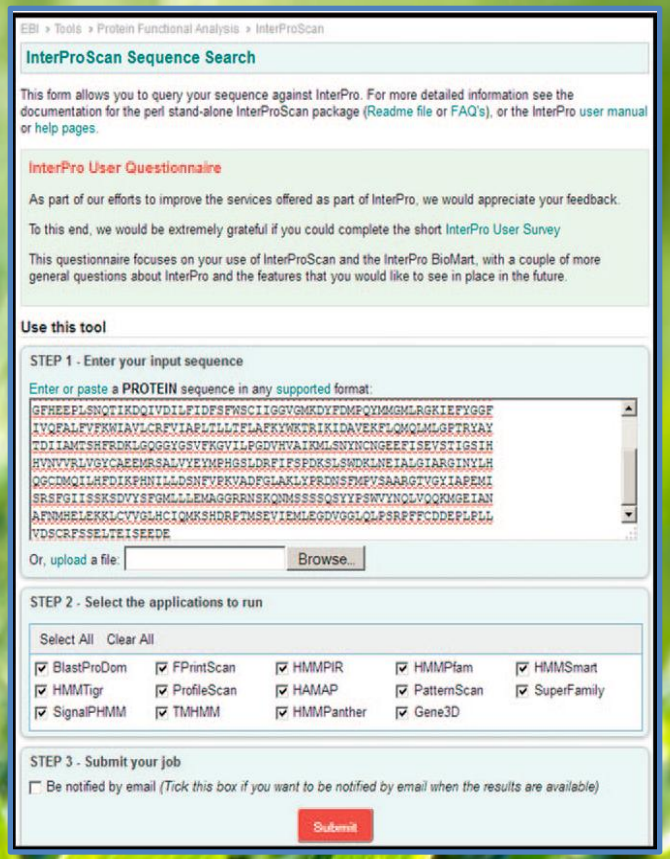
Library of Congress Control Number: 2013936960

Functional Annotation of Genomes Using an Integrative Approach.

Discovering protein signatures using Inter-Pro.

❖Inter-Pro (an Integrated source of Protein Domains and Functional Sites) is a collaborative project which provides an integrated layer upon the most commonly used suites of signature databases (Pfam, PROSITE, PRINTS,ProDom, SMART, TIGRGAMMs).

❖InterProScan, a tool available from InterPro, can be run on a web-based interface (http://ebi.ac.uk/interpro/scan.html) or via a local installation. The current web version requires a protein sequence as the input file.



❖On the web interface InterProScan will return the results in a graphical “Picture View” format and can be viewed or retrieved in Table View, Raw Output, or XPML output.

❖ It is important to note that InterProScan calculates a checksum for your query sequence and uses it to look up precomputed tables to find an equivalent signature (that equals that checksum).

❖For bulk data sets such as in a whole genome, a standalone downloadable version is available from ftp.ebi.ac.uk/pub/databases/interpro/iprscan/.

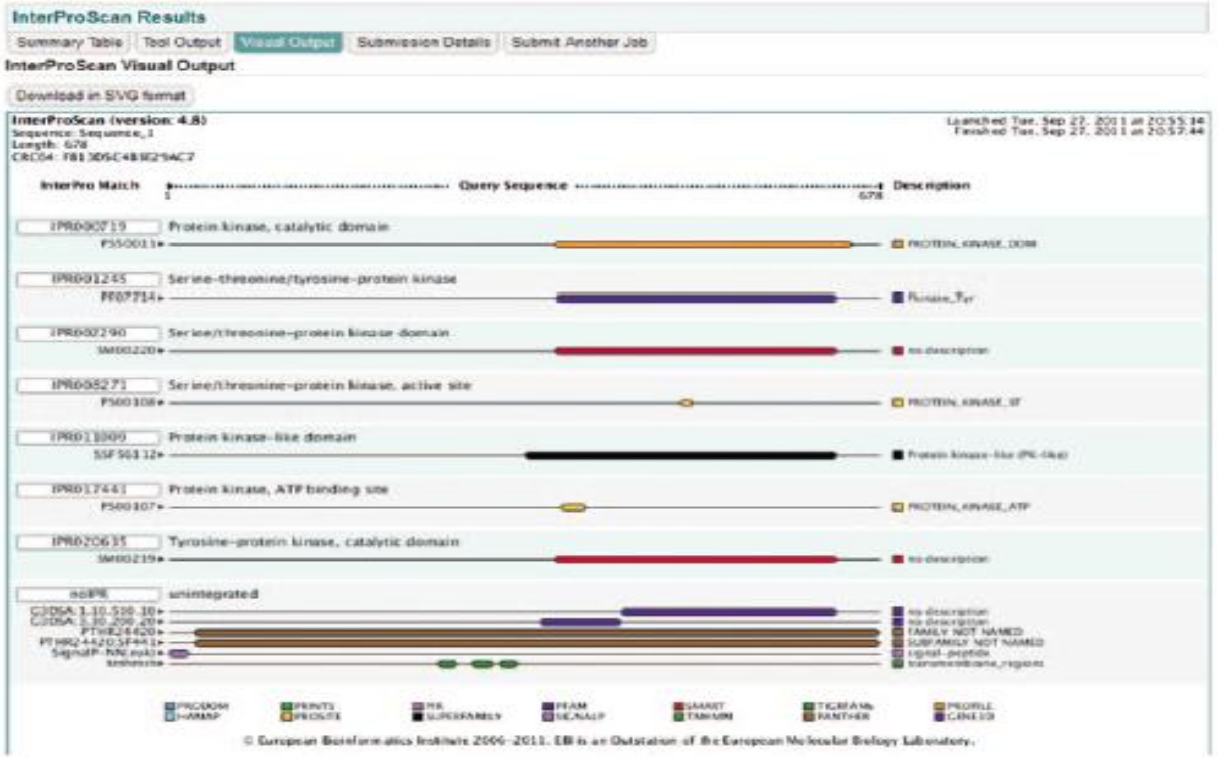


Fig. 7.3 The graphical output from the web interface version of the InterProScan for the rice protein LOC_Os01g02350.1: InterPro entries (number on the left-hand side of the image) and InterPro name (right-hand side of the image) is returned for seven different applications (domain databases). PANTHER, GENE3D,SIGNALP and TMHMM are not included in the integrated InterProScan and therefore no InterPro entries are returned for these applications, but if the corresponding domains are recognized they are marked on the sequence (*bottom* of the window)

SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
NATURAL VARIATION IN PLANTS
AGRICULTURAL GENETICS

Received
26 June 2014

Accepted
13 February 2015

Published
19 March 2015

A combinatorial approach of comprehensive QTL-based comparative genome mapping and transcript profiling identified a seed weight-regulating candidate gene in chickpea

Deepak Bajaj[1], Hari D. Upadhyaya[2], Yusuf Khan[1], Shouvik Das[1], Saurabh Badoni[1], Tanima Shree[1], Vinod Kumar[3], Shailesh Tripathi[4], C. L. L. Gowda[2], Sube Singh[2], Shivali Sharma[2], Akhilesh K. Tyagi[1], Debasis Chattopdhyay[1] & Swarup K. Parida[1]

[1]National Institute of Plant Genome Research (NIPGR), Aruna Asaf Ali Marg, New Delhi 110067, India, [2]International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru 502324, Telangana, India, [3]National Research Centre on Plant Biotechnology (NRCPB), New Delhi 110012, India, [4]Division of Genetics, Indian Agricultural Research Institute (IARI), New Delhi 110012, India.

Correspondence and requests for materials should be addressed to S.K.P. (swarup@nipgr.

RESULTS

This study was undertaken to validate and genotype genome-wide physically mapped 1632 Single Nucleotide Polymorphisms and 500 single sequence repeats markers showing in-silico polymorphism between **ICC4958 (desi)** and **ICC12968 (kabuli)** using **Illumina Golden Gate Assay, Gel based Assay and fluorescent dye labeled automated fragment analyser**

The marker genotyping and robust field phenotyping information of mapping individuals were used to develop high resolution intra-specific genetic linkage map for identification of major QTLs associated with pod and seed number/plant and 100 seed weight in chickpea.

One of 126.8kb major genomic region harbouring a strong SW- associated robust **QTL(Caq'SW1.1:169.1-171.3cM)** has been delineated by integrating high-resolution QTL mapping with comprehensive marker-based comparative genome mapping and differential expression profiling. This identified one potential regulatory SNP in the cis-acting element of candidate ERF(ethylene responsive factor) TF(transcription factor) gene governing seed weight in chickpea.

Table 2 | Significant QTLs associated with pod and seed number/plant and seed weight identified and mapped on eight chickpea LGs/chromosomes using an intra-specific mapping population (ICC 4958 × ICC 12968)

				2012			2013		
QTLs	LGs/chromosomes	Marker intervals with genetic positions (cM)	Markers associated with QTLs	LOD	PVE (R^2%)	A	LOD	PVE (R^2%)	A
Caq'PN1.1 & *Caq'SN1.1*	CaLG(Chr01)	Ca-IISNP18 (16.8) to Ca-IISNP20 (18.6)	Ca-IISNP18	NS	NS	NS	4.6	12.7	8.9
Caq'PN1.2 & *Caq'SN1.2*	CaLG(Chr01)	Ca-IISNP32 (29.5) to Ca-IISNP35 (32.1)	Ca-IISNP34	5.1	7.8	4.5	NS	NS	NS
Caq'PN1.3 & *Caq'SN1.3*	CaLG(Chr01)	Ca-IISNP116 (127.6) to Ca-IISNP119 (130.9)	Ca-IISNP116	8.5	12.4	6.3	7.8	14.7	5.4
Caq'PN2.1 & *Caq'SN2.2*	CaLG(Chr02)	Ca-IISNP330 (154.9) to Ca-IISNP332 (157.5)	Ca-IISNP331	10.5	10.8	11.4	9.8	12.4	8.9
Caq'PN3.1 & *Caq'SN3.1*	CaLG(Chr03)	Ca-IISNP398 (24.7) to Ca-IISNP401 (27.7)	Ca-IISNP399	6.4	9.4	10.9	NS	NS	NS
Caq'PN4.1 & *Caq'SN4.1*	CaLG(Chr04)	Ca-IISNP649 (12.7) to Ca-IISNP651 (123.1)	Ca-IISNP649	11.4	19.8	12.9	10.2	18.5	9.5
Caq'PN5.1 & *Caq'SN5.1*	CaLG(Chr05)	Ca-IISNP831 (89.1) to Ca-IISNP834 (92.5)	Ca-IISNP832	8.5	12.5	6.5	7.3	14.7	5.7
Caq'PN6.1 & *Caq'SN6.1*	CaLG(Chr06)	Ca-IISNP1104 (154.8) to Ca-IISNP1108 (157.6)	Ca-IISNP1106	NS	NS	NS	4.8	8.5	5.1
Caq'PN7.1 & *Caq'SN7.1*	CaLG(Chr07)	Ca-IISNP1332 (199.5) to Ca-IISNP1335 (202.2)	Ca-IISNP1335	5.7	6.8	10.2	NS	NS	NS
[a]*Caq'PN8.1* & *Caq'SN8.1*	CaLG(Chr08)	Ca-IISNP1487 (152.9) to Ca-IISNP1489 (154.6)	Ca-IISNP1487	9.6	9.5	3.1	8.5	8.8	2.5
Caq'SW1.1	CaLG(Chr01)	Ca-IISNP151 (169.1) to Ca-IISNP154 (171.3)	Ca-IISNP152	13.8	25.8	10.7	12.6	24.5	8.9
Caq'SW1.2	CaLG(Chr01)	Ca-IISSR55 (196.5) to Ca-IISSR25 (199.5)	Ca-IISSR55	8.5	11.6	3.9	NS	NS	NS
Caq'SW1.3	CaLG(Chr01)	Ca-IISNP186 (202.6) to Ca-IISSR56 (204.9)	Ca-IISNP186	9.5	13.4	5.1	8.7	16.3	4.6
[a]*Caq'SW2.1*	CaLG(Chr02)	Ca-IISNP234 (44.2) to Ca-IISNP238 (47.6)	Ca-IISNP236	7.6	9.7	4.7	7.0	9.5	5.1
Caq'SW3.1	CaLG(Chr03)	Ca-IISNP465 (115.6) to Ca-IISNP468 (118.5)	Ca-IISNP466	NS	NS	NS	7.2	10.6	8.6
Caq'SW4.1	CaLG(Chr04)	Ca-IISNP579 (3.3) to Ca-IISNP581 (5.8)	Ca-IISNP579	8.5	10.3	4.7	7.8	11.4	3.8
[a]*Caq'SW4.2*	CaLG(Chr04)	Ca-IISNP749 (275.3) to Ca-IISNP752 (278.8)	Ca-IISNP751	6.9	9.8	7.1	6.5	8.5	6.7
Caq'SW6.1	CaLG(Chr06)	Ca-IISNP972 (13.2) to Ca-IISNP975 (16.0)	Ca-IISNP974	4.7	8.7	8.9	NS	NS	NS

*_Caq'PN1.1_ (_Cicer arietinum_ QTL for pod number on chromosome 1 number 1), _Caq'SN1.2_ (_Cicer arietinum_ QTL for seed number on chromosome 1 number 2) and _Caq'SW2.1_ (_Cicer arietinum_ QTL for 100-seed weight on chromosome 2 number 1), PVE: Percentage of phenotypic variation explained by QTLs, A: Additive effect; positive additive effect infers alleles from ICC 4958 with increasing trait values. Details regarding Ca-IISSR and Ca-IISNP markers are provided in the Supplementary Table S1.

[a]known QTLs from previous studies by Cobos et al. (28,44), Hossain et al. (45), Varshney et al. (22) and Gowda et al. (37). NS: non-significant QTLs.

Contents lists available at SciVerse ScienceDirect

Biotechnology Advances

journal homepage: www.elsevier.com/locate/biotechadv

Research review paper

Rice functional genomics research: Progress and implications for crop genetic improvement

Yunhe Jiang, Zhaoxia Cai, Weibo Xie, Tuan Long, Huihui Yu, Qifa Zhang *

National Key Laboratory of Crop Genetic Improvement and National Center of Plant Gene Research, Huazhong Agricultural University, Wuhan 430070, China

ARTICLE INFO

Available online 24 August 2011

Keywords:
Oryza sativa
Functional genomic resources
Gene cloning
Rice 2020
Green Super Rice

ABSTRACT

Rice is a staple food crop and has become a reference of monocot plant for functional genomic research. With the availability of high quality rice genome sequence, there has been rapid accumulation of functional genomic resources, including: large mutant libraries by T-DNA insertion, transposon tagging, and chemical mutagenesis; global expression profiles of the genes in the entire life cycle of rice growth and development; full-length cDNAs for both *indica* and *japonica* rice; sequences from resequencing large numbers of diverse germplasm accessions. Such resource development has greatly accelerated gene cloning. By the end of 2010, over 600 genes had been cloned using various methods. Many of the genes control agriculturally useful traits such as yield, grain quality, resistances to biotic and abiotic stresses, and nutrient-use efficiency, thus have potential utility in crop genetic improvement. This review was aimed to provide a comprehensive summary of such progress. We also presented our perspective for future studies.

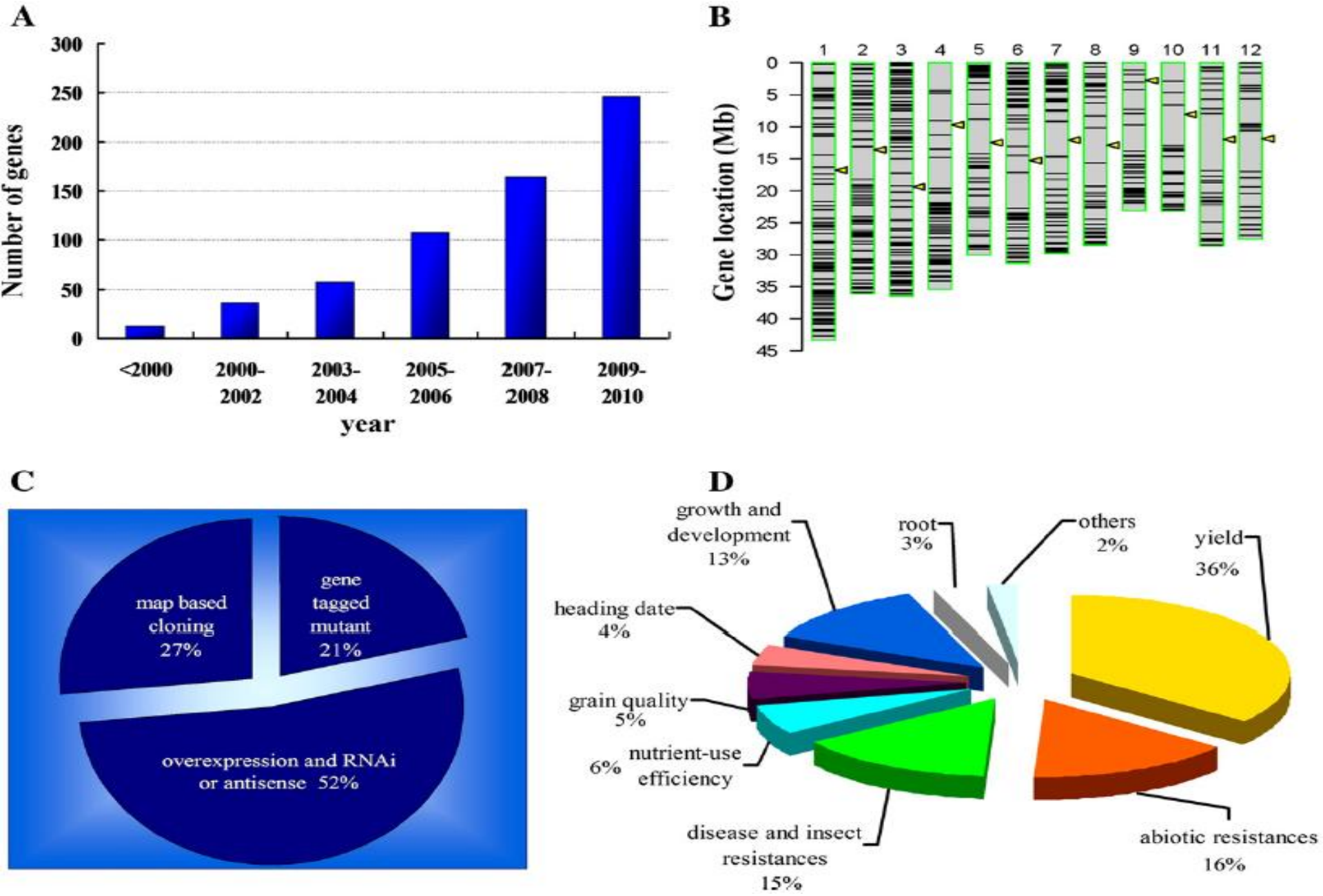


Fig. 1. Current status of isolated genes in rice. (A) Number of genes cloned in different years. (B) Chromosome localizations of 592 isolated genes. Each thin horizontal line represents a gene. The yellow triangles indicate the positions of the centromeres. (C) Numbers of genes isolated by different methods. (D) Functional categories of the isolated genes.

Biotechnology Advances 31 (2013) 1308–1318

Contents lists available at ScienceDirect

Biotechnology Advances

journal homepage: www.elsevier.com/locate/biotechadv

Research review paper

Genomics-based precision breeding approaches to improve drought tolerance in rice

B.P. Mallikarjuna Swamy, Arvind Kumar *

International Rice Research Institute, DAPO Box 7777, Metro Manila, Philippines

ARTICLE INFO

Article history:
Received 7 December 2012
Received in revised form 23 April 2013
Accepted 8 May 2013
Available online 20 May 2013

Keywords:
Rice
Drought
Grain yield
Genetics
Genomics
Transcriptomics
Proteomics
QTL
Allele mining
Comparative genomics

ABSTRACT

Rice (*Oryza sativa* L.), the major staple food crop of the world, faces a severe threat from widespread drought. The development of drought-tolerant rice varieties is considered a feasible option to counteract drought stress. The screening of rice germplasm under drought and its characterization at the morphological, genetic, and molecular levels revealed the existence of genetic variation for drought tolerance within the rice gene pool. The improvements made in managed drought screening and selection for grain yield under drought have significantly contributed to progress in drought breeding programs. The availability of rice genome sequence information, genome-wide molecular markers, and low-cost genotyping platforms now makes it possible to routinely apply marker-assisted breeding approaches to improve grain yield under drought. Grain yield QTLs with a large and consistent effect under drought have been indentified and successfully pyramided in popular rice mega-varieties. Various rice functional genomics resources, databases, tools, and recent advances in "-omics" are facilitating the characterization of genes and pathways involved in drought tolerance, providing the basis for candidate gene identification and allele mining. The transgenic approach is successful in generating drought tolerance in rice under controlled conditions, but field-level testing is necessary. Genomics-assisted drought breeding approaches hold great promise, but a well-planned integration with standardized phenotyping is highly essential to exploit their full potential.

Genomics of Drought Tolerance

In rice, several efforts were made to understand the differential expression of genes under drought stress by transcription profiling using RT-PCR and Microarray Techniques. Kathiresan *et al* (2006) carried out Microarray analysis of rice panicles selected from a drought-stressed plant and reported up-regulation of several stress induced genes such as GTP- binding protein 3, cellulose synthase-6, heat shock cognate potein, DNA repair protein, reductase, zinc finger protein, actin depolymerising factor and pectin estrase. Rabbani *et al* (2003) identified 62 drought inducible genes in 2 week old drought stressed rice seedling

Popular rice land race N22 (Nagina 22) is a natural choice for much molecular characterization work because of its drought tolerance. A drought-induced cDNA library was constructed from N22 and trascription profiling was performed, 589 putative stress related genes were identified, and using this information, candidate gene were deduced for drought QTLs.

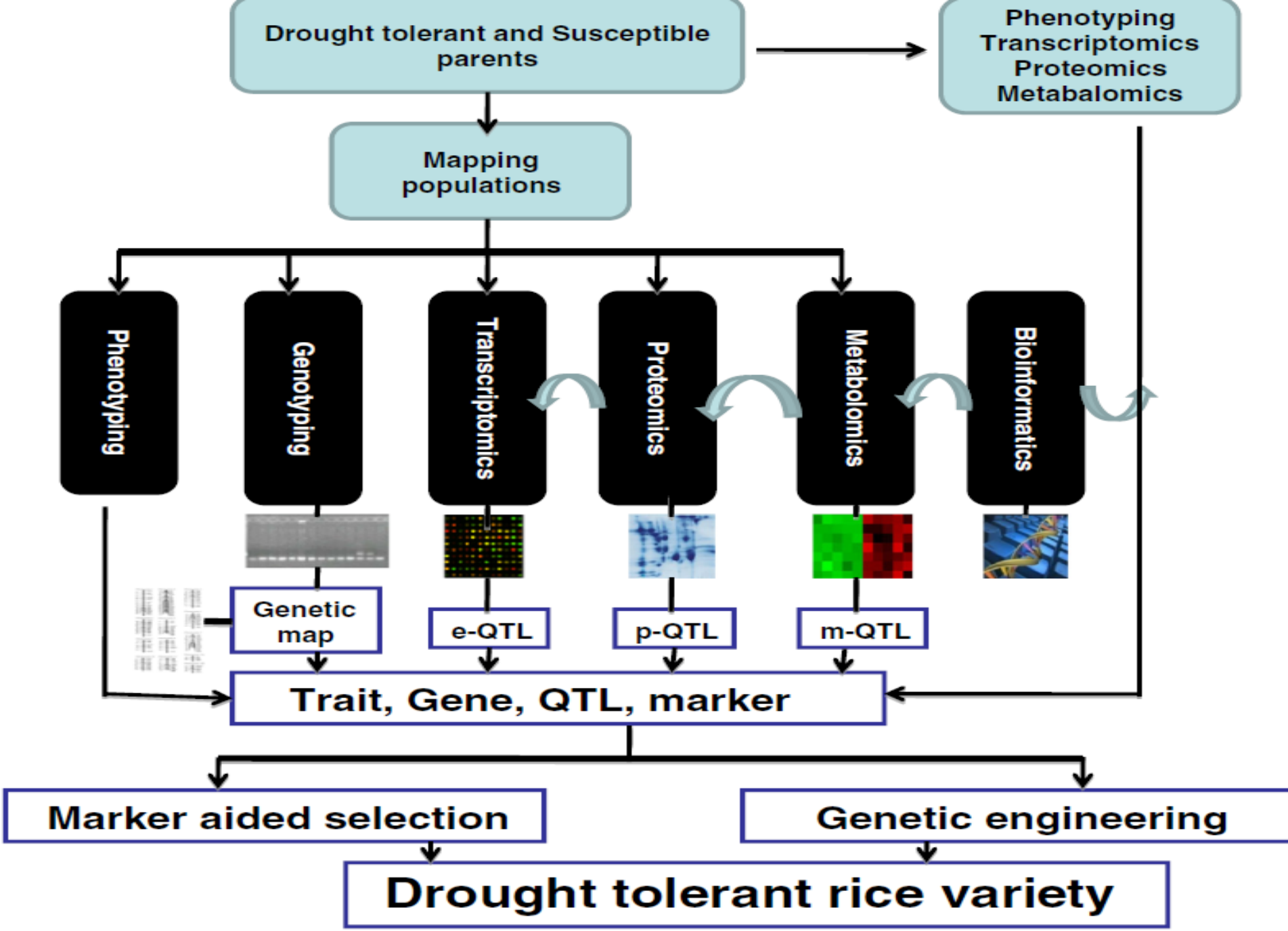


Fig. 1. Genomics assisted breeding approaches to improve drought tolerance in rice.

Table 2
Major effect QTL grain yield under drought.
Source: IRRI.

Variety	QTL	Donor	Ecosystem	Additive effect	Phenotypic variance (R2)
Vandana	$qDTY_{12.1}$	Way Rarem	Upland	47	33
Way Rarem	$qDTY_{2.3}$	Vandana	Upland, lowland	15	6
	$qDTY_{3.2}$	Vandana	Upland, lowland	23	16
IR64	$qDTY_{1.1}$	N22	Lowland	24	17
	$qDTY_{2.2}$	Adaysel	Lowland	14	6
	$qDTY_{4.1}$	Adaysel	Lowland	6	11
	$qDTY_{9.1}$	Adaysel	Lowland	29	19
	$qDTY_{10.1}$	Adaysel	Lowland	18	17
Swarna	$qDTY_{1.1}$	N22	Lowland	29	13
	$qDTY_{1.1}$	Dhagaddeshi	Lowland	25	32
	$qDTY_{2.1}$	Apo	Lowland	23	7
	$qDTY_{3.1}$	Apo	Lowland	30	27
MTU1010	$qDTY_{1.1}$	N22	Lowland	16	13
Basmati334	$qDTY_{8.1}$	Swarna	Lowland	23	7
Sabitri	$qDTY_{12.1}$	IR74371-46-1-1	Lowland	25	47
	$qDTY_{3.2}$	IR77298-5-6-18	Lowland	16	19
Samba Mahsuri	$qDTY1_{1.1}$	IR55419-04	Lowland	32	14
TDK1	$qDTY_{3.1}$	IR55419-04	Lowland	10	7
	$qDTY_{6.1}$	IR55419-04	Lowland	12	9
	$qDTY_{6.2}$	IR55419-04	Lowland	16	10

ICRISAT

International Crops Research Institute for the Semi-Arid Tropics

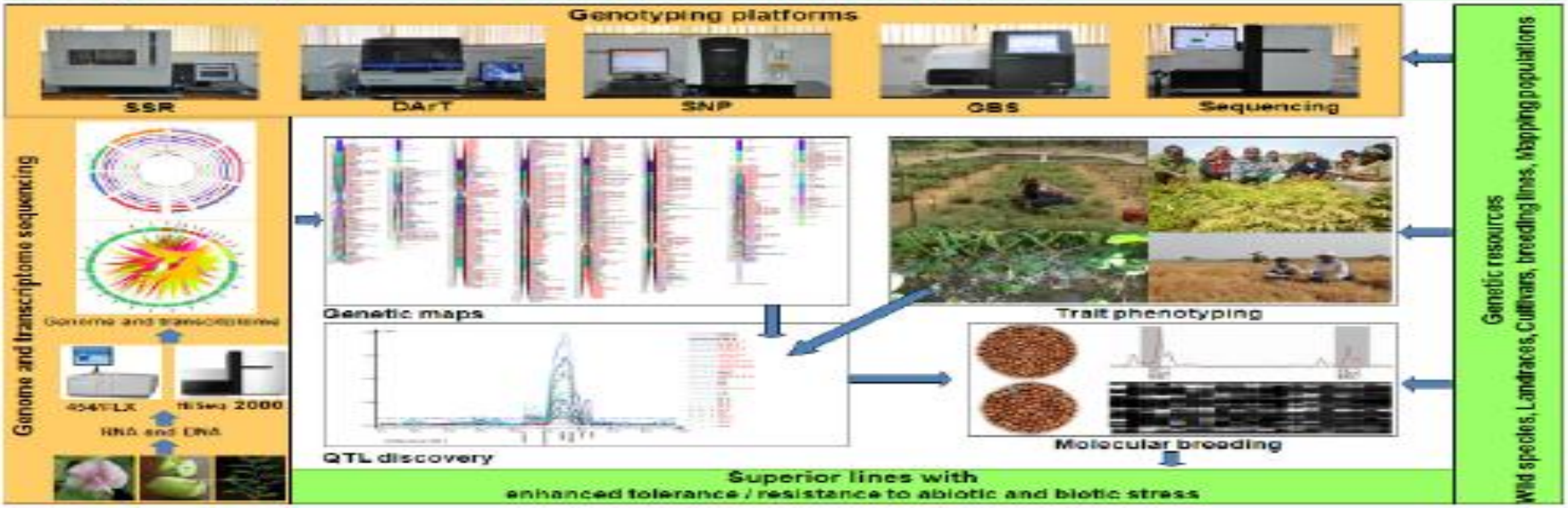
Center of Excellence in Genomics

Striving towards efficient breeding and research

About the CEG

Efficient plant breeding requires high-throughput allele determination at low cost for better prediction of an individual's phenotype from its genotype. This is the primary reason for the establishment of ICRISAT's Center of Excellence in Genomics (CEG). To cater the needs of molecular breeding community, the CEG has three main components i.e. applied genomics research and sequencing, high quality marker genotyping services and capacity building in modern genomics and molecular breeding.

Integrated genomics and breeding activities



http://ceg.icrisat.org

V NGGIBCI-2015

5th International Conference
on
Next Generation Genomics and Integrated Breeding for Crop Improvement

ICRISAT, Patancheru, India

ICRISAT
Science with a human face

TheScientist
EXPLORING LIFE, INSPIRING INNOVATION

Dr. Rajeev K Varshney

Principal Scientist-Applied Genomics
Research Program Director - Grain Legumes
Director-Center of Excellence in Genomics
Winthrop Research Professor-The University of Western Australia
International Crops Research Institute for the Semi-Arid Tropics
Patancheru
India
r.k.varshney@cgiar.org

Dr. Rajeev Varshney, an Indian national and Principal Scientist is serving ICRISAT as a Research Program Director, Grain Legumes and Director - Center of Excellence in Genomics. In addition to serving ICRISAT, Rajeev, in his dual appointment earlier served CGIAR Generation Challenge Program based in Mexico as Theme Leader for six years. Before joining ICRISAT, he worked at Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Germany, for five years.

Rajeev has a basic background in molecular genetics and possess more than 15 years research experience in international agriculture. The primary contribution of Rajeev Varshney includes genome sequencing of pigeonpea, chickpea, peanut, pearl millet, sesame, mung bean and azukibean and the first generation of molecular breeding products in chickpea and groundnut, in addition to large-scale genomic resources such as molecular markers, transcriptome assemblies, high density genetic maps and QTLs for a range of traits in legumes. Rajeev has a prolific publication record with >200 publications in leading journals of international reputes including Nature, Nature Biotechnology, Nature Communications, PNAS etc., 10 edited books and special issues (as Guest Editor) for several journals to his credit. He has been a frequent invited speaker in several national/ international conferences including G-8 Conference on "Open Data for Agriculture", FAO conference on "Application of Biotechnologies in Developing Countries" and brainstorming session on digital agriculture chaired by Mr Bill Gates. Rajeev has won several awards/fellowships including Elected Fellow of Indian National Science Academy (INSA), National Academy of Agricultural Sciences (NAAS), India; Crop Science Young Scientist Award (Crop Science Society of America); INSA Young Scientist Medal; Associate NAAS Fellow; NASI Young Scientist Platinum Jubilee Award, and The Greater Good Initiative Award of Illumina.

Next Generation Genomics and Integrated Breeding in Legumes

Rajeev K Varshney

International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru -502324, Hyderabad, India
School of Plant Biology and Institute of Agriculture, The University of Western Australia, Australia
Email: r.k.varshney@cgiar.org

Food security is a major global concern as still >800 million people are suffering from starvation and malnutrition in Sub-Saharan Africa (SSA) and Asia. Legume crops such as chickpea, pigeonpea and groundnut play important role to ensure food and nutritional security in developing countries. For developing climate change-ready legume varieties with enhanced yield and nutritional quality, next generation genomics and integrated breeding approaches are being used. In this context, large-scale genomic resources including draft genome sequences, germplasm re-sequencing, transcriptome assemblies, molecular markers, high density genetic maps and cost-effective genotyping platforms have been developed in these legume crops. In parallel, specialized genetic stocks such as reference set, RIL, MAGIC and NAM population have been developed. These germplasm resources are being genotyped using high-density genotyping/re-sequencing and phenotyped in different agro-climatic conditions. Analysis of these massive genotyping and phenotyping data is providing markers/genes associated with traits of interest using linkage mapping and genome wide association studies (GWAS). Efforts have also been initiated to deploy genomics-assisted breeding (GAB) approaches and superior lines with enhanced drought tolerance in chickpea and disease resistance in chickpea and groundnut have been developed. However, it is essential to empower NARS partners for sustainable deployment of GAB to feed ever-growing population in developing countries of SSA and Asia.

TheScientist
EXPLORING LIFE, INSPIRING INNOVATION

Dr Frank Ordon

JKI-Federal Research Centre for Cultivated Plants
Institute for Resistance Research and Stress Tolerance
Quedlinburg
Germany
frank.ordon@jki.bund.de

Dr. Frank Ordon is head of the Institute for Resistance Research and Stress Tolerance of the Julius Kühn-Institute (JKI), Federal Research Centre for Cultivated Plants, Germany, and honorary professor for 'molecular resistance breeding' at the Martin-Luther-University, Halle-Wittenberg, Germany. He studied agricultural science at the Justus Liebig University in Giessen, Germany, where he also got his PhD and state doctorate. He is the editor in chief of *Plant Breeding* and is a member of several editorial boards, e.g. *Theoretical and Applied Genetics*, and *Journal of Applied Genetics*, and of scientific advisory boards, e.g. of the IPK Gatersleben. He is chair of the Research Committee of the Wheat Initiative. Frank has a basic background in classical and molecular plant breeding, with special emphasis on breeding for resistance against viral and fungal pathogens in barley and wheat.

His primary contributions include genetic analyses of resistance, and the development of molecular markers for major resistance genes and QTL, especially against virus diseases, up to gene isolation. Besides this, he is working on improving tolerance to abiotic stress in several crop species. Frank has published the results of his studies in more than 100 papers in peer-reviewed journals.

Genomics Based Breeding Research for Improving Resistance to Biotic and Abiotic Stress in Cereals

Perovic D[1], Serfling A[1],Perner K[1], Mitterbauer E[1], Knöchel N[1], Fettköther T[1], Wehner G[1], Lehnert H[1], Silvar C[3], Krämer I[1], Habekuss A[1], Kopahnke D[1], Graner A[2], Stein N[2], Ordon F[1*]

[1]Julius Kühn-Institute (JKI), Institute for Resistance Research and Stress Tolerance, Erwin-Baur-Str. 27, 06484 Quedlinburg, Germany
[2]Leibniz Institute for Genetics and Crop Plant Research (IPK), Corrensstr. 3, 06466 Gatersleben, Germany
[3]Department of Ecology, Plant and Animal Biology, University of Coruña, Campus da Zapateira s/n, 15071A Coruña, Spain
*Email: frank.ordon@jki.bund.de

Feeding the earth´s growing population against the background of climate change is one of the major future challenges. In this respect wheat and barley are of special importance. However, both are hit by many fungal and viral pathogens causing severe yield losses. In addition, yield is negatively affected by abiotic stress, e.g. drought. Therefore, breeding research on improving resistance/tolerance to biotic stress and on adaptation to climate change is a prerequisite for facing future challenges.

Today genomic resources like the Infiniumi Select genotyping arrays, the Genome Zipper, comprising a virtual linear order of genes of different monocot species, next generation sequencing techniques, e.g. genotyping by sequencing (GBS), exome capture, or RNAseq and MACE, allow efficient marker development and gene identification in bi-parental populations as well as by genome wide association studies (GWAS). This tool box facilitates (a) a more efficient exploitation of genetic resources, (ii) an efficient marker development for genes and QTL involved in resistance/tolerance to biotic and abiotic stress, (iii) an enhanced identification of candidate genes followed by gene isolation, (iv) allele mining and allele based breeding, (v) a faster implementation of new breeding goals. Examples for genomics-based breeding research for improving resistance to biotic stress, e.g. on a *P. triticina* resistance derived from *T. monococcum*, to abiotic stress, e.g. drought, as well as for the implementation of new breeding goals, e.g. adaptation to rising CO_2 concentrations are given.

TheScientist
EXPLORING LIFE, INSPIRING INNOVATION

Invited Speaker Bio

Dr Trilochan Mohapatra

Director
ICAR-Central Rice Research Institute
Cuttack
India
tmnrcpb@gmail.com, directorcrri@sify.com

Dr. Trilochan Mohapatra is currently the Director of the Central Rice Research Institute, Cuttack, Odisha, India. Prior to this, he worked as a researcher and teacher for about 20 years at the National Research Centre on Plant Biotechnology, Indian Agricultural Research Institute (IARI), New Delhi. He has been working in the area of molecular genetics and genomics. Dr. Mohapatra has over 135 research papers in national and international journals of repute and several book chapters. His research accomplishments include the development of the first high-yielding Basmati rice variety resistant to bacterial leaf blight through molecular marker assisted selection, and physical mapping and genome sequencing of rice and tomato. He has received several honours and awards in recognition of his excellent academic and research contributions, including the INSA Young Scientist Award, Prof. LSS Kumar Memorial Award, NAAS-Tata Award, IARI BP Pal Award, DBT Bio-science Award and NASI-Reliance Industries Platinum Jubilee Award. He is a Fellow of the Indian National Science Academy, New Delhi, National Academy of Sciences-India, Allahabad, and the National Academy of Agricultural Sciences, New Delhi.

Genome-Wide Association of SNPs in Stress Responsive Genes with Salinity Tolerance in Rice

Vinod Kumar[1], Anshuman Singh[1], S.V. Amitha Mithra[1], Swarup K. Parida[1], S.L. Krishnamurthy[2], Sourabh Jain[1], Kapil K. Tiwari[1], Pankaj Kumar[1], A.R. Rao[3], S.K. Sharma[2], J.P. Khurana[4], N.K. Singh[1], T. Mohapatra[1,5,*]

[1]National Research Centre on Plant Biotechnology, New Delhi 110012, India; [2]Central Soil Salinity Research Institute, Karnal 132001, India; [3]Indian Agricultural Statistics Research Institute, New Delhi 110012, India; [4]University of Delhi, South Campus, New Delhi 110021, India; [5]Central Rice Research Institute, Cuttack, Odisha 753006, India
*Email: tmnrcpb@gmail.com

Rice growth is affected due to salinity stress, to a varying degree at all stages starting from germination through maturation. Reduction in major yield components, including tiller and spikelet numbers, has been reported to be the major cause of yield loss (27-50%) in rice during early reproductive panicle initiation stage under salinity stress. In Asia, 12 million ha of land area is thought to be salinity affected, with India having more than 50% of this area. We implemented global transcriptome profiling of control and salt stress (200 mM NaCl) seedlings of two rice genotypes (CSR30 and Taraori Basmati) using the next-generation Illumina Solexa Genome Analyzer. We identified 288 and 140 differentially expressed salinity-tolerant and sensitive transcripts/genes in CSR30 and Taraori Basmati, respectively. Functional annotation of these transcripts corresponded with different known (19%) and candidate (67%) salinity tolerance and unknown expressed (14%) rice genes. Pathway analysis enabled us to localize 128 known and candidate salinity tolerant genes in various known salinity stress-responsive pathways, including salt overly sensitive (SOS) and calcium signalling. Integration of differential expression profiling of genes with that of whole genome re-sequencing data enabled identification of single nucleotide polymorphism (SNP) in the salinity stress-responsive rice genes. In an alternative approach, the SNP information was utilized to carry out genome wide association studies (GWAS) to identify loci controlling salinity tolerance. A custom designed SNP array based on 6000 stress-responsive genes, distributed at an average physical interval of <100 kb on 12 rice chromosomes was used to genotype 220 rice accessions using Infinium assay. Genetic association was analysed with 12 different traits recorded on these accessions under field conditions at reproductive stage. We identified 20 SNPs (loci) significantly associated with Na^+/K^+ ratio, and 44 SNPs with other traits observed under stress condition. In addition to a known quantitative trait locus (QTL) for Na^+/K^+ ratio under stress near *Saltol* locus on chromosome 1, we found GWAS peaks representing new QTLs on chromosomes 4, 6 and 7. The current association mapping panel contained mostly *indica* accessions which can serve as source of novel salt tolerance genes and alleles. The study helped in unveiling genomic regions/candidate genes regulating salinity stress tolerance in rice.

Chloroplast Targeting: Potential Tool for Functional Genomics in Pulses

Poornima KN*, **Das A, Singh NP**

Division of Plant Biotechnology, ICAR-Indian Institute of Pulses Research, Kanpur-208 024, Uttar Pradesh, India
*Email: poornimakn04@gmail.com

Genes require different *cis*-acting elements for their expression and localisation. Chloroplast has been a very important organelle constituting many photosynthetic proteins. Majority of these proteins are encoded by the nucleus and then transported into the chloroplast. The chloroplast transport is aided by a *cis*-element called the chloroplast Transit Peptide (cTP) that localises all the essential proteins into chloroplast. The present study focuses on isolation of genes encoding chloroplast targeting proteins, including transit peptide sequences from chickpea and pigeonpea genome. All the transit peptide sequences were aligned, their conserved motifs identified, and their phylogeny deciphered. Sequence analysis revealed several signature consensus sequences that marks chloroplast targeting, such as the conserved 'homology block' Gly-X-Arg-XXX-Val and the presence of valine and alanine at -3 and -1 sites of all the analysed peptides. The current study will help understand the structural and functional aspects of chloroplast targeting in pulses. Further, foreign proteins expressed at very low levels or toxic to the cytoplasm can be localised to intracellular compartments of chloroplast by fusing with cTP. It can also facilitate the genome editing of chloroplast genome, and also for targeting of different cellular proteins into the chloroplast with higher and stable expression.

Accelerating Precision and Efficiency of Breeding Programs through Center of Excellence in Genomics (CEG)

Integrated genomics and breeding activities

Genotyping

Crops

- Chickpea
- Pigeonpea
- Groundnut
- Sorghum
- Pearl millet
- Finger millet
- Rice
- Wheat
- Maize
- Barley
- Oats
- Sweet potato
- Tomato
- Potato
- Onion
- Garlic
- Cassava
- Tobacco
- Cotton
- Mango
- Guava
- Litchi
- Banana
- Mulberry

Projects

- Fingerprinting of cultivars
- Genetic diversity analysis
- Linkage mapping
- Association mapping
- QTL mapping
- Marker-assisted selection
- Genomic selection

Genomics

Genome sequences

- Pigeonpea - 72.7% coverage *(Nature Biotech 2012, 30:83-89)*
- Chickpea - 73.8% coverage *(Nature Biotech 2013, 31:240-246)*
- Sorghum - 94.5% coverage (available through US-led team)
- Groundnut (Announced on 2 April 2014)
- >1000 legume genomes sequenced
- Pearl millet (in progress)

Marker resources

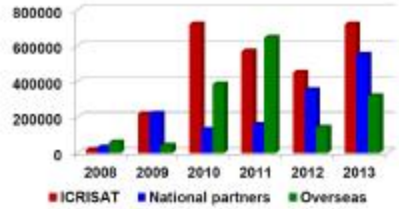
- >10,000 SSRs across mandate crops
- >10,000 SNPs across mandate crops
- High density DArT arrays in all mandate crops
- KASPar, Illumina GoldenGate and VeraCode assays for chickpea, pigeonpea and groundnut

Capacity building

Students from sub-Saharan Africa and South Asia

Beneficiaries

- NARS partners in Africa, Asia, Europe and Latin America
- Agricultural research institutes
- Research foundations
- Universities
- Small-scale breeding companies



Overview on data generated for SSR, DArT and SNP markers

ICRISAT
Science with a human face

International Crops Research Institute for the Semi-Arid Tropics

WHERE ARE WE HEADING?

- Although gene function prediction has been an active area of research for the past 15 years, its use in plant science has been limited.
- To exploit this technology we need :

✓ more data.

✓ better assessment of data and tool quality.

✓ easy access to the data and tools.

✓ high-throughput experimental validation.

- To reach the goal of breeding better plants for the future we have a long way to go.

THERE IS SO MUCH MORE TO BE DISCOVERED

THANK YOU...