

TransPlant User Needs Survey

Context: Following EU review recommendations, this survey aims at collecting bioinformatics stakeholders needs and to describe some existing initiatives launched in the field of agronomical research to be compared to TransPlant scope. The goal is to identify potential needs not already covered by TransPlant project. The second objective is to draw the landscape of possibly overlapping initiative in the field or related fields to better coordinate developments and possibly avoid redundancy.

Based to this survey and on other actions of WP2 and WP4, EU will be given in summer 2013 a review on existing initiatives in the field of transnational research for bioinformatics in crop science and an overview of the needs of the stakeholders in terms of bio-informatic applications.

This survey contains 41 questions. Thanks you in advance for your involvement.

1. Personal information : Family name (optional)

2. First name (optional)

3. Name of your Institute, company...

4. Category of your provenance

	large	medium	small
commercial	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
university	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
non university academic	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
academic policy-related	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

*5. Position

- Graduate student
- Professor
- Research scientist

Other (please indicate)

6. How old are you ?

- <30 31-40 41-50 51-60 >60

TransPlant User Needs Survey

*7. What are your fields of expertise?

- Structural genomics
- Functional genomics
- Quantitative genetics
- Genetics and applied mathematics and statistics
- Population genetics
- Breeding
- Agronomy
- Physiology
- Plant Biology
- Biochemistry
- Bioinformatics

Other (please specify)

*8. What data are you interested in ?

please rank according to the importance of the data category for your research

	1 : maximum priority	2 : medium priority	3: low priority	4: no priority
genome sequences (NGS included)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
sequence variation (SNP, CNV, PAV...)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
genetics maps (markers)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
genetics maps (QTLs)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
curated gene models	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
metabolic profiles	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
protein profiles	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
resources for natural genetic diversity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
segregating populations or linkage disequilibrium panels	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
phenotypes (classification, ontologies)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
phenotypes (experimental measurements)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
data citation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
long term preservation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Other (please specify)

TransPlant User Needs Survey

*9. What analysis do you do ?

- Genome structure and evolution
- Comparative genomics
- Genetic diversity / Population genetic
- Genotype to phenotype analysis (GWAS, QTL detection...)
- Breeding
- Map based cloning
- Developmental biology
- Gene/protein/metabolite network inferences
- Modeling of plant development and adaptation (to check with ecophysio)

Other (please specify)

TransPlant User Needs Survey

10. How much storage do you need to maintain your projects ? (please indicate the size and the unit: Mb, Tb, Pbytes...)

***11. Do you have problems in accessing sufficient storage for your needs, and if so, why ?**

You need to choose at least one choice

insufficient local
infrastructure (due to high
throughput scale)

Insufficient human
resources to manage/
develop/maintain a local
infrastructure

Other (please specify)

12. Do you have unmet needs in bioinformatic analysis ?

13. Do you submit your data to public repositories?

please give the name(s) of the databases (international, consortium database, repositories, local database) and the type of data that you put inside

name of the database

type of data that you put
inside

***14. What data types do you submit?**

reads

images

pre-computed data from analysis software tools (blast results, varscan results for example)

Other (please specify)

TransPlant User Needs Survey

*15. Are these data derived from ?

- Your own experiments produced in your lab
- Your own experiments produced by external platforms or production centers
- Public data

Other (please specify)

*16. What types of elaborated (compute and expertized) data do you want to share ?

- Genome sequences, assembled sequenced
- Resequencing data: sequence variation (SNPs, CNV, structural variants)
- Genotyping data (alleles, haplotypes, localisation on genomes, frequency)
- Annotation data (gene models, gene function prediction, gene ontologies)
- Genetic mapping data (maps, markers, QTLs, metaQTLs)
- Exome sequencing data
- Phenotyping data (traits, phenotypes, experimental conditions)
- Association studies data (traits, phenotypes, markers effects, statistical values, LD)
- Genomic selection data
- Expression data (RNASeq)
- Expression data (Arrays)
- Methylation data (ChipSeq)
- Repeats data (transposable elements)
- Small RNA
- Orthologous gene comparisons

Other (please specify)

TransPlant User Needs Survey

*17. What types of data do you want to access ?

- gene annotations and variation data
- gene annotations and genetic mapping (markers, QTL, metaQTLs) data
- genes annotations and association studies data (effects of SNP markers on traits)
- gene annotations with all objects mapped around: variations, markers, QTL, expression, repeats...
- genotyping and phénotyping data
- gene annotations and gene networks
- gene annotations, protein and metabolites
- gene orthologous species comparison
- system biology

Other (please specify)

18. Please specify your preferred relevance criteria to sort results from database queries

- age
- taxonomy
- data source (for integrated search)
- keywords in hit description
- manual screening of results
- authors

Other (please specify)

TransPlant User Needs Survey

19. What types of data should be released publicly in standard data formats?

- contigs
- chromosomes
- primers
- cDNA
- statistical parameters values from experimental or analysis software
- SNP markers
- genotyping values
- phenotyping values
- maps

Other (please specify)

20. Which data formats do you currently use for sharing/uploading analysing data? (GFF3, VCF, OWL, XML, tablesheets...)

21. What types of tools for visualization and browsing are you using : please indicate for each tool: its name, its provider, the URL web to access it (if exists), if it is a public or private resource

22. Do you need additional tools for query and if yes, which tools or functionalities are missing today ?

TransPlant User Needs Survey

23. According to your answer to the previous question, are your needs concerning :

- A specific domain that is insufficiently well developed ?
- a problem of data access ?
- a problem of data integration ?

Other (please specify)

24. What is your preferred way to get informed about updates and new to databases, tools and services ?

- passive - check by my own
- social media (e.g. facebook, twitter, google+)
- mailing lists
- science newsletter
- RSS feed

Other (please specify)

25. What would your dream bioinformatics resource allow you to do that you cannot do at the moment?

26. Do you need to be able to upload data to analyse against the public data sets?

- yes no

27. Do you need to be able to co-analyse private (confidential) and public data?

- yes no

*28. Do you need compute resources to be provided to you ?

- yes no

TransPlant User Needs Survey

29. How do you normally access compute resources ?

- Central power unit (CPU) on a cluster ?
- CPU on a private Cloud ?
- CPU on a public Cloud ?

Other (please specify)

* 30. What types of softwares do you usually use to analyze your data ? please specify for each tool: the source of these software (opensource, company, local tools..) and the category (Assembly, gene annotation, repeats annotations, QTL analysis, GWAS analysis, Arrays analysis, RNASeq analysis, ChipSeq Analysis, proteomics analysis, modelisation analysis..., Galaxy tools, workflows...

31. What types of software do you think, is missing today to analyse your data ? please indicate the category (see previous questions to help or add new categories

* 32. Do you have difficulties when analysing your data ?

- yes no

33. If yes, why is it difficult ?

- Little relevant data is available about my species of interest
- Because of your species complexity
- Poor or inadequate software documentation
- Limited information about strengths/weaknesses of available software
- No bioinformatics/statistics support available
- Problems in hiring appropriately skilled staff
- Large scale data (algorithm or tool not well adapted)
- Rapid technological change

Other (please specify)

34. Are you involved in projects in which bioinformatics developments are ongoing/planned?

- yes no

35. Give the name of the project, its funding (if it is not confidential)

36. Do you need help in understanding software developments ?

yes

no

37. If yes, please indicate to do what

38. Do you need help in bioanalysis ?

yes

no

39. if yes: to analyse what ?

40. The transPLANT project is attempting to coordinate activities in plant bioinformatics in Europe. What activities would you like the project to undertake to support the community?

- A unique web portal to get news, to get list or resources and tools, to access to integrated data
- A forum of discussion and exchange
- A annual or binannual meeting
- User training sessions (software, data access, analysis,...)
- Newsletters, one or twice a year

Other (please specify)

41. Do you have any comments or recommandation to add, helping bioinformatics communities to fit your needs in an optimal way ?