

Sébastien CARRERE, Emmanuel COURCELLE, Marion VERDNAUD, Eric BIOT, Erika SALLET, Emeline DELEURY, Loïck LEDANTEC, Cécile FIZAMES, Jean-Pierre GAUTHIER, Vincent SAVOIS, Susete ALVES-CARVALHO, Philippe GREVET, Véronique BRUNAUD, Fabrice LEGEAL, Bernhard GSCHLOESSL, Virginie GARCIA and Jérôme GOUZY.
<http://cati-bbric.toulouse.inra.fr>

Principe

Nous présentons une application d'architecture orientée service (SOA) avec pour application l'analyse RNA-seq.

Au travers d'une application web, l'utilisateur accède de façon transparente aux données et méthodes d'analyse disponibles sur un réseau constitué de 8 laboratoires.

Ce réseau distribué met à disposition de la communauté des données d'expression relatives à des espèces d'intérêt Agronomique [3,4] et ainsi permet l'identification de transcrits différentiellement exprimés.

Les données et méthodes d'analyses sont distribuées sur le réseau et la communication se fait via des web-services BioMOBY [1] enregistrés dans un annuaire central.

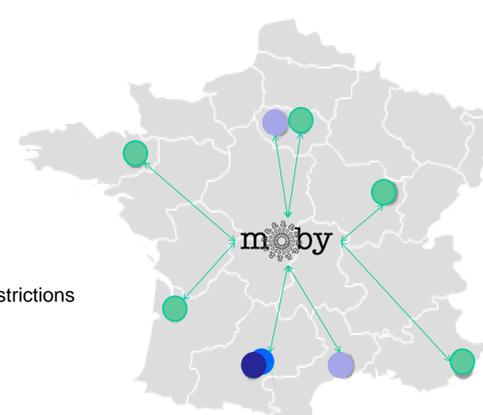
La technologie des web-services permet l'interopérabilité avec des systèmes externes (ex: Gene Atlas)

L'architecture orientée service appliquée à l'analyse d'expression différentielle offre une grande flexibilité et réactivité.

En effet, dès qu'un fournisseur dépose de nouvelles données sur son serveur, celles-ci sont disponibles à l'ensemble du réseau (aux restrictions d'accès près).

De même, si une nouvelle méthode d'analyse est implémentée sur un site, elle est utilisable immédiatement par les utilisateurs du système.

Afin de garantir une qualité de service maximum, l'ensemble du réseau est surveillé tant au niveau matériel qu'au niveau fonctionnel. Par soucis de transparence, cette surveillance est directement accessible à tout utilisateur sur le portail de l'application.



- BioMOBY central
- Analyses DGE
- Données
- Application RNA-seq
- Données
- Analyses DGE [5]
- Analyse TopGO
- Serveur externe

Fonctionnement

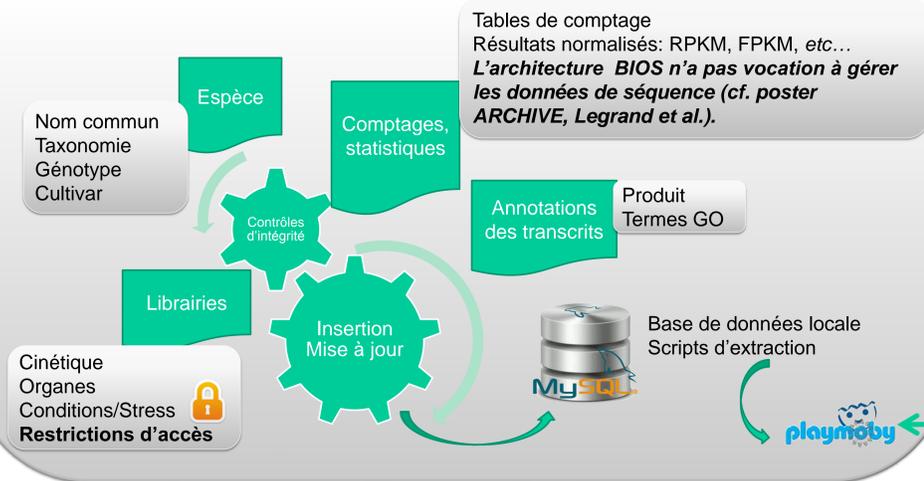
Mettre à disposition de nouvelles données

L'ajout ou la mise à jour de données sur un serveur du réseau se fait via des scripts Perl et des fichiers de configuration et/ou tabulés.

Ces outils ainsi que les scripts d'extraction et leur interface BioMOBY sont déployés lors de l'installation du nouveau serveur.

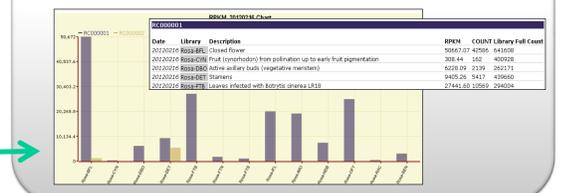
Mettre à disposition une nouvelle méthode d'analyse

Le bioinformaticien écrit un wrapper Perl de sa nouvelle méthode et le déploie sur le réseau via le framework PlayMOBY [2] installé lors de la mise en service de son serveur



Accès programmatique (API BioMOBY)

Ex: « GeneAtlas »



Une application de l'architecture BIOS à l'analyse d'expression différentielle

L'utilisateur sélectionne l'espèce.
La liste des espèces dépend du niveau d'authentification

Species selection interface showing a list of species and their associated data.

L'utilisateur sélectionne la méthode d'analyse et les filtres associés

Analysis method and filter selection interface.

L'utilisateur sélectionne la liste des librairies à comparer.

Library selection interface for comparison.

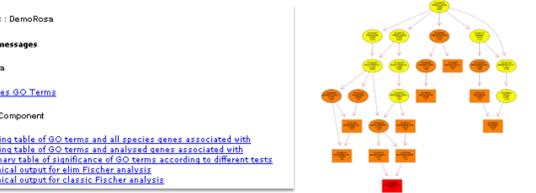
L'utilisateur fixe les filtres sur les transcrits à analyser

Transcript filter setting interface.

RNA-seq analysis: DemoRosa

System messages, Pools composition, Raw data filters, Statistical Method & Filters, Results.

| Accession | length | type | version | product | adjust | chisprob | f_hvo |
|-----------|--------|------|----------|---|--------|----------|------------------------|
| R0001240 | 1470 | mRNA | 20090808 | PR002228 Alcohol dehydrogenase, zinc-containing, conserved site, PR013149 Alcohol dehydrogenase, zinc-binding | 0 | 0 | 0.0012135906884722 |
| R0004112 | 1187 | mRNA | 20090808 | PR016293 Peptidase M1DA, matrix metalloproteinase | 0 | 0 | 0.000112172154230886 |
| R0000001 | 1310 | mRNA | 20090808 | PR000001 | 0 | 0 | 0.022472707839461 |
| R0002887 | 723 | mRNA | 20090808 | PR002987 | 0 | 0 | 0.000623311863702472 |
| R0037739 | 451 | mRNA | 20090808 | PR000726 Glycoside hydrolase, family 15, catalytic | 0 | 0 | 0.00000100130419871883 |
| R0001205 | 927 | mRNA | 20090808 | PR001087 Lipase, ODSL | 0 | 0 | 0.0001947536686508113 |
| R0000038 | 1662 | mRNA | 20090808 | PR003593 ATPase, AAA-type, core, PR003959 ATPase, AAA-type, conserved site | 0 | 0 | 0.00049849490961978 |
| R0007283 | 328 | mRNA | 20090808 | PR010800 Glycine rich | 0 | 0 | 0.00000750979149039123 |
| R0000279 | 2254 | mRNA | 20090808 | PR002355 Multicopper oxidase, copper-binding site, PR014707 Coagulation factor VII | 0 | 0 | 0.0000035045649551591 |



Un réseau de confiance transparent

Surveillance des serveurs

| Host | Status | Last Check | Duration | Status Information |
|-----------------|--------|---------------------|-----------------|--|
| bordeaux_bios | UP | 2013-06-19 17:32:05 | 49d 21h 51m 46s | PING OK - Packet loss = 0%, RTA = 4.00 ms |
| bordeaux_bios | UP | 2013-06-19 17:31:55 | 49d 21h 51m 36s | PING OK - Packet loss = 0%, RTA = 20.00 ms |
| evry_bios | UP | 2013-06-19 17:31:55 | 49d 21h 51m 36s | PING OK - Packet loss = 0%, RTA = 16.00 ms |
| rennes_bios | UP | 2013-06-19 17:31:55 | 49d 21h 51m 36s | PING OK - Packet loss = 0%, RTA = 12.00 ms |
| sophia_bios | UP | 2013-06-19 17:31:55 | 49d 21h 51m 36s | PING OK - Packet loss = 0%, RTA = 12.00 ms |
| versailles_bios | UP | 2013-06-19 17:31:55 | 49d 21h 51m 36s | PING OK - Packet loss = 0%, RTA = 4.00 ms |
| dijon_bios | UP | 2013-06-19 17:33:05 | 89d 6h 34m 1s | PING OK - Packet loss = 0%, RTA = 0.00 ms |



Tests fonctionnels des web-services quotidiens

| Service | Date | Execution time |
|-----------------------------|---------------------|-----------------------|
| BIOSGetAssemblyVersions | 2013.06.19 02:00:04 | Execution time: 2 s. |
| BIOSGetGoTermsFromSpecies | 2013.06.19 02:00:25 | Execution time: 19 s. |
| BIOSGetLibrariesFromSpecies | 2013.06.19 02:00:27 | Execution time: 2 s. |
| BIOSGetNodeLead | 2013.06.19 02:00:28 | Execution time: 1 s. |

Redondance des web-services d'analyse, équilibrage de charge

| CPUs | Number | Load | | | Memory | |
|-------------------------|--------|-------|-------|--------|-----------|-----------------|
| | | 1 min | 5 min | 15 min | Total | Physical Memory |
| bios.dijon.inra.fr | 16 | 1.10 | 1.04 | 1.01 | 128000216 | 565196 |
| bios.versailles.inra.fr | 4 | 0.01 | 0.00 | 0.00 | 16777432 | 15897836 |
| bios.rennes.inra.fr | 6 | 0.00 | 0.00 | 0.00 | 23593176 | 222016 |
| bios.sophia.inra.fr | 4 | 0.00 | 0.00 | 0.00 | 33554648 | 1797404 |
| bios.toulouse.inra.fr | 4 | 1.12 | 1.06 | 1.01 | 33554648 | 13194628 |
| bios.evry.inra.fr | 8 | 0.02 | 0.02 | 0.00 | 71303384 | 9546760 |
| bios.supagro.inra.fr | 4 | 0.01 | 0.00 | 0.00 | 16777432 | 14904128 |
| bios.bordeaux.inra.fr | 6 | 1.02 | 1.02 | 1.00 | 25166040 | 233436 |

Références

- Wilkinson, M.D., Senger, M., Kawas, E., Bruskiwich, R., Gouzy, J., Noïrot, C., Bardou, P., Ng, A., Haase, D., Saiz, E. de, et al. (2008) Interoperability with Moby 1.0—it's better than sharing your toothbrush! *Briefings in bioinformatics*, 9, 220-31
- Bertrand Néron, Hervé Ménager, Corinne Maurais, Nicolas Joly, Julien Maupetit, Sébastien Carrere, Pierre Tuffery, and Catherine Letondal (2009) Moby: a new full web bioinformatics framework. *Bioinformatics* 25: 3005-3011.
- Arnick Dubois, Sébastien Carrere, Olivier Raymond, Benjamin Pourreau, Ludovic Cottret, Aymeric Roccia, Jean-Paul Onesto, Soulaïman Sakr, Rossitza Atanassova, Sylvie Baudino, Fabrice Foucher, Manuel Brts, Jérôme Gouzy, Mohammed Bendahmane (2012) Transcriptome database resource and gene expression atlas for the rose. *BMC Genomics*, 13, 638
- Deleury E, Dubreuil G, Elarogovan N, Wajnberg E, Reichhart JM, Gourbal B, Duval D, Baron O, Gouzy J, Coustau C. (2012) Specific versus non-specific immune responses in an invertebrate species evidenced by a comparative de novo sequencing study. *PLoS One*, 7(3)
- Anders, S. and Huber, W. (2010) Differential expression analysis for sequence count data. *Genome biology*, 11, R106