MICROSCOPE

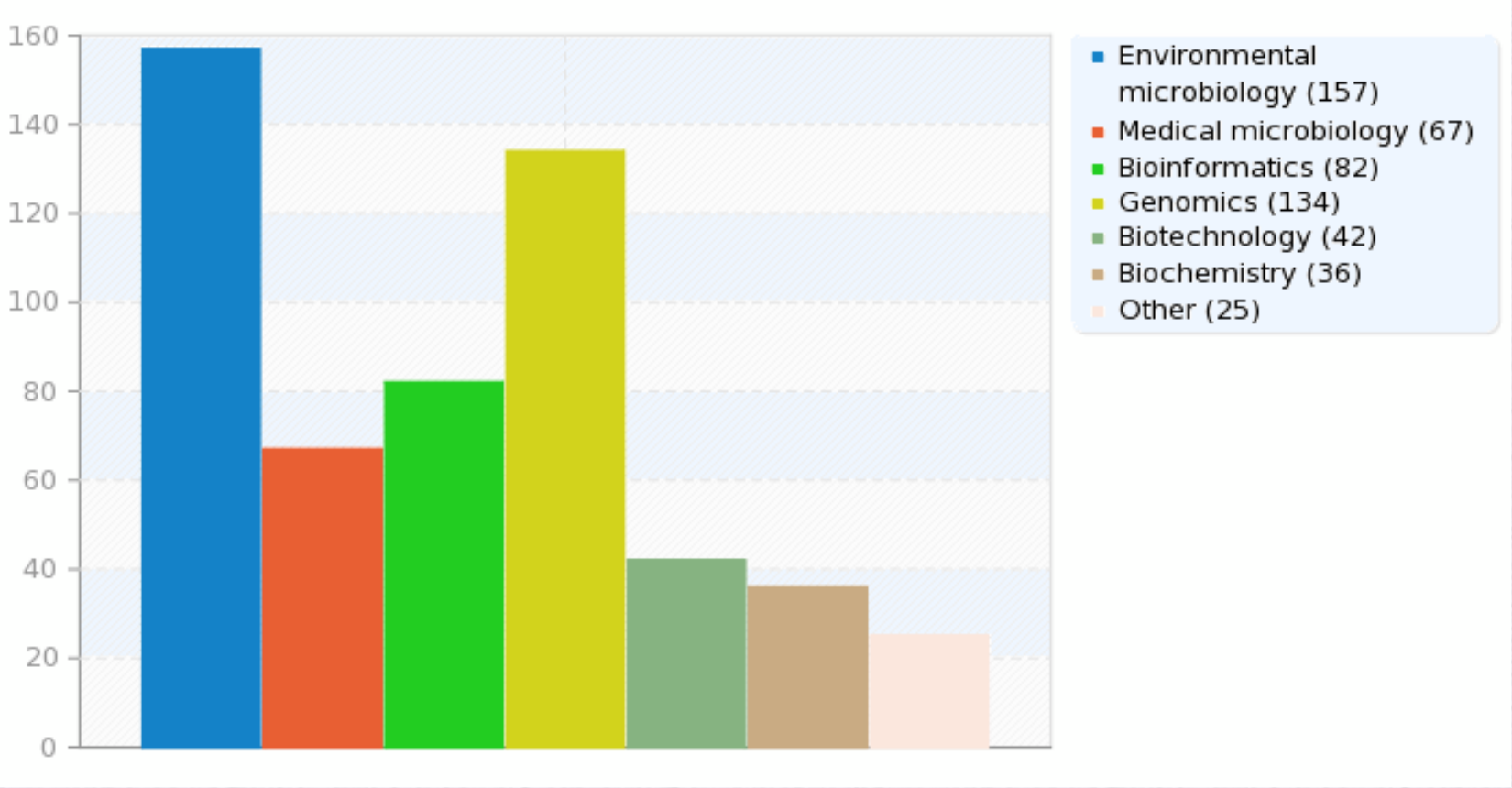
Sastisfactory survey 2014

Participation

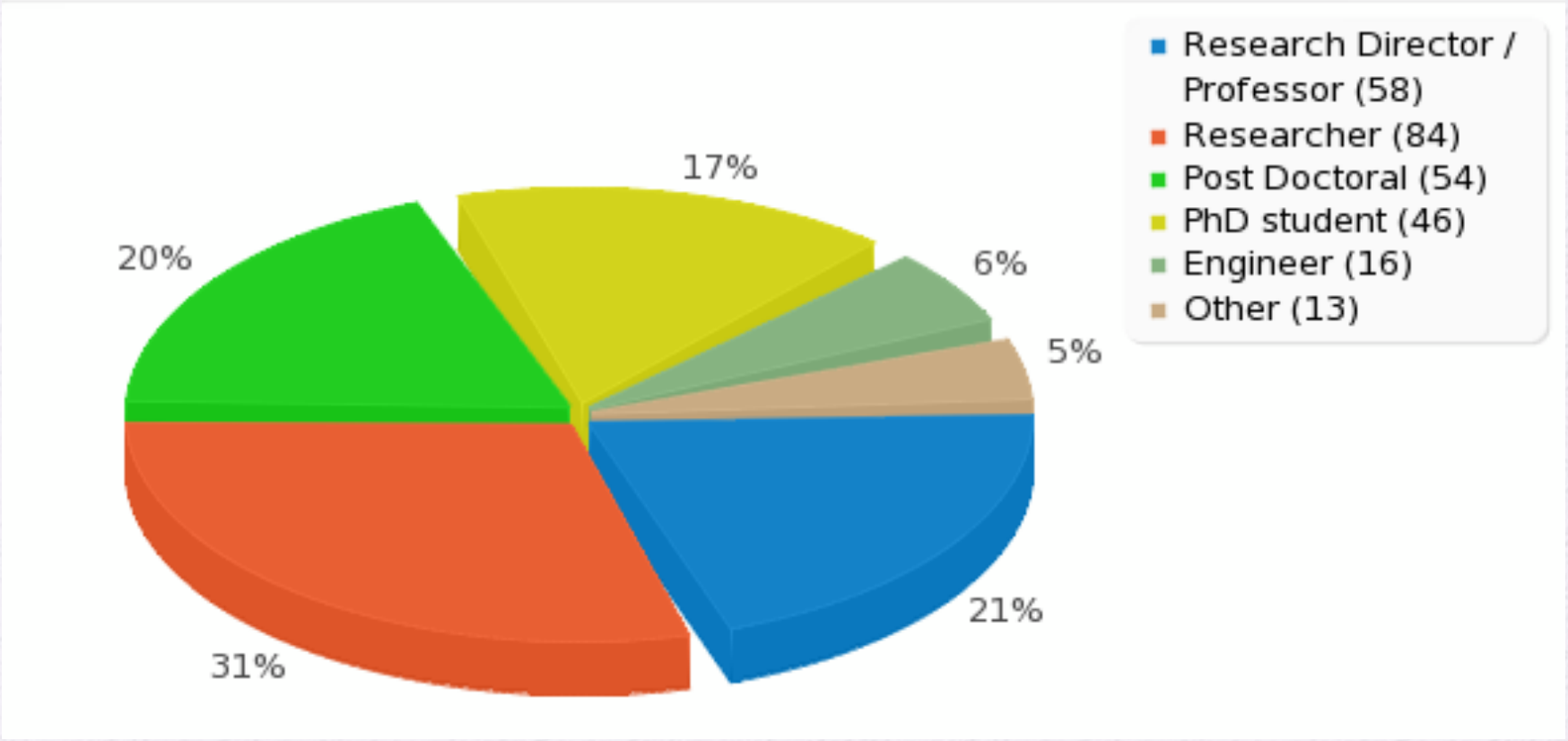
- 2298 user accounts / 327 « active » accounts
  (account number with at least one connection per month; average on 2014)
- 271 answers
  - 12% (compared to total accounts)
  - 83% (compared to active accounts)
- Many comments and suggestions

Who are you ?

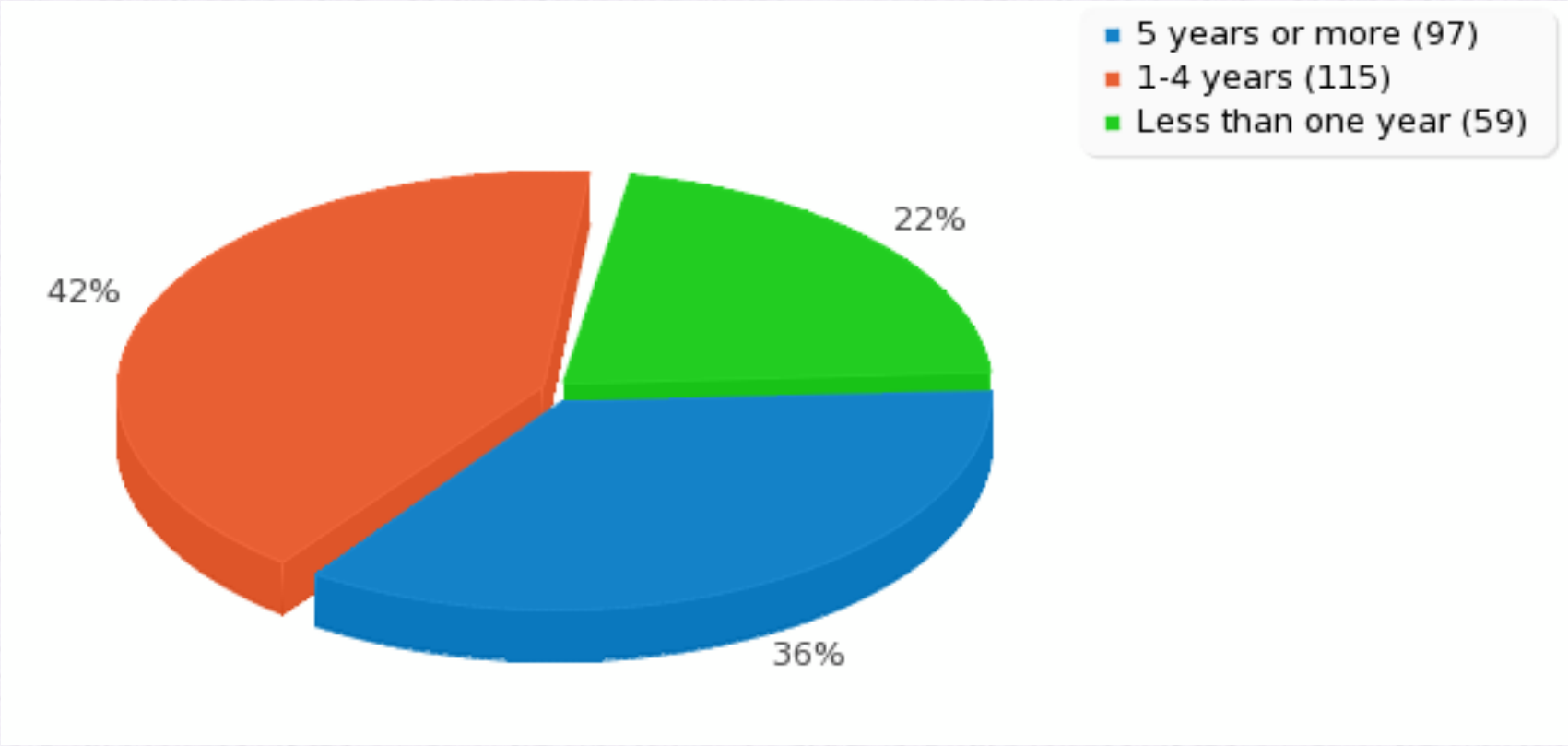
Research field / area



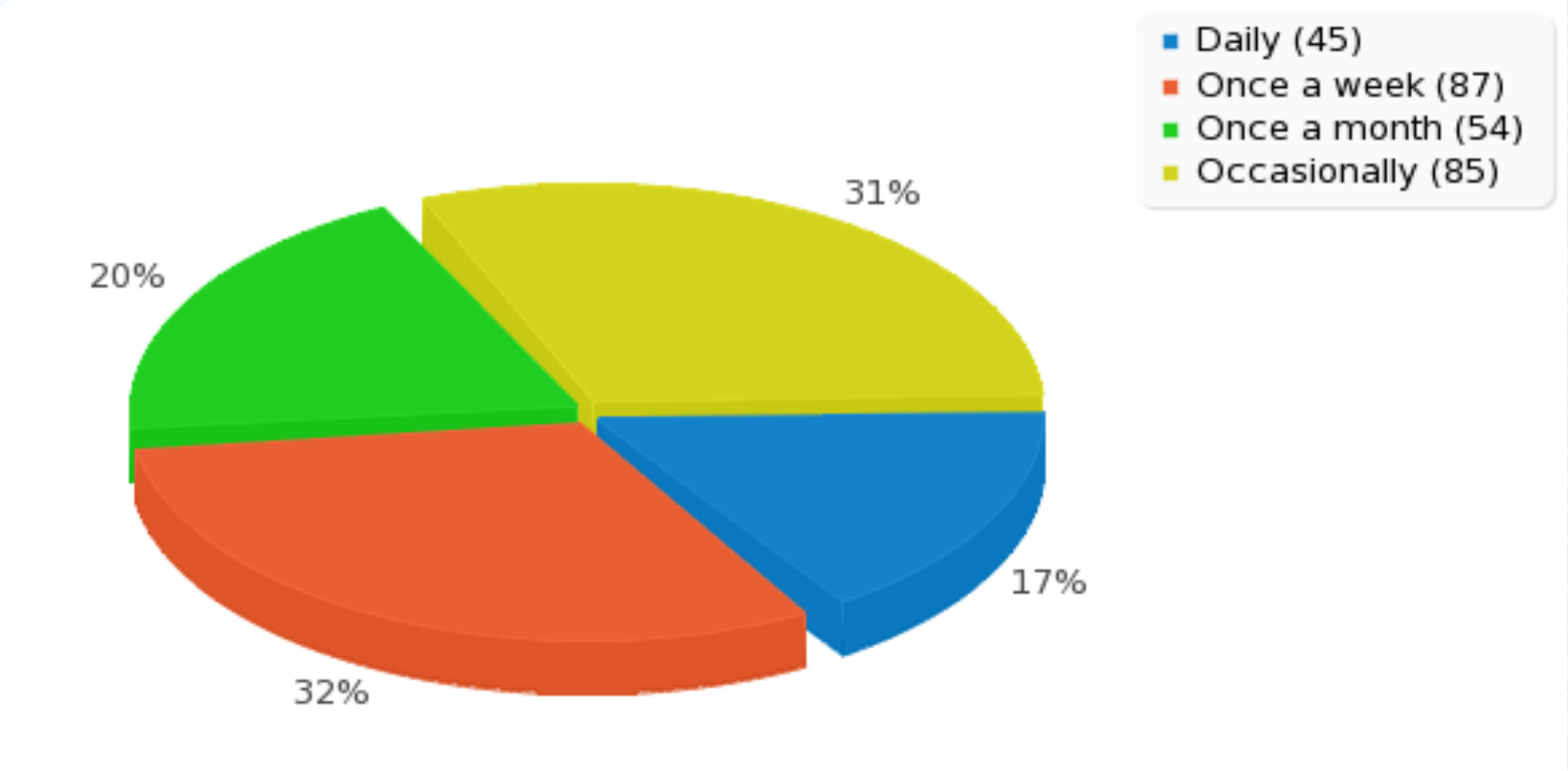
Current position



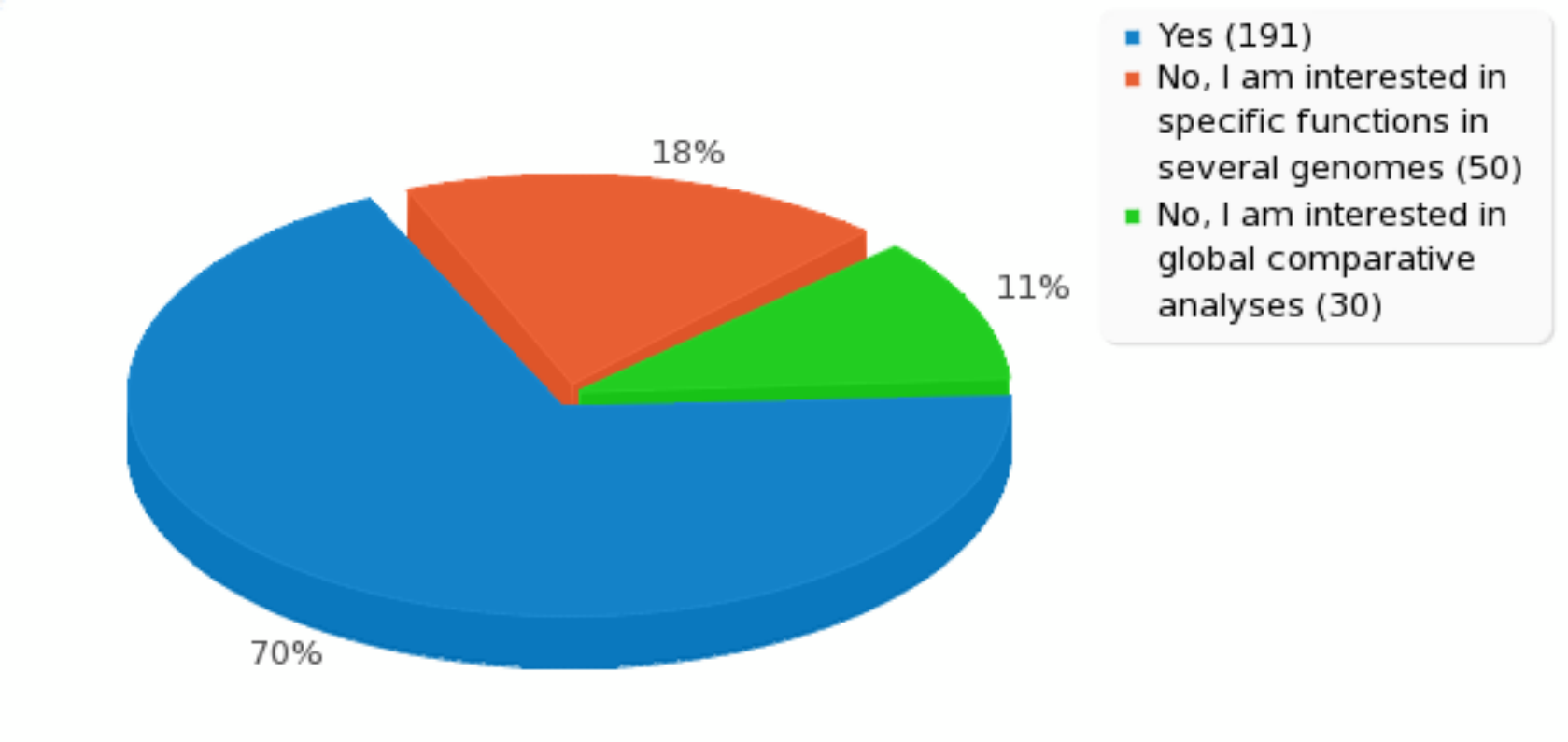
How long have you been using the MicroScope platform ?



How often do you use the platform ?

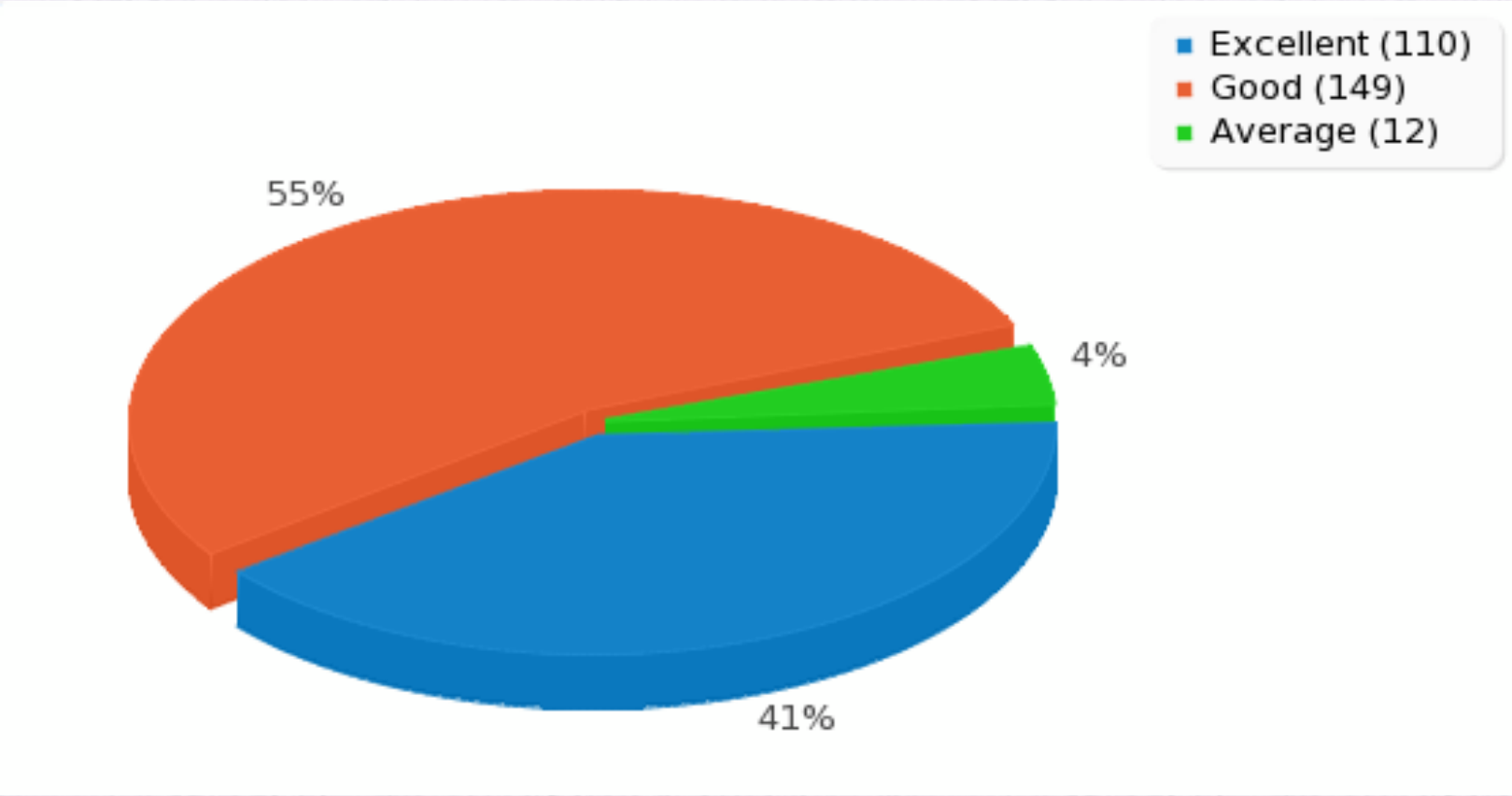


Do you use MicroScope for the analysis of specific specie(s)?



General use of MicroScope

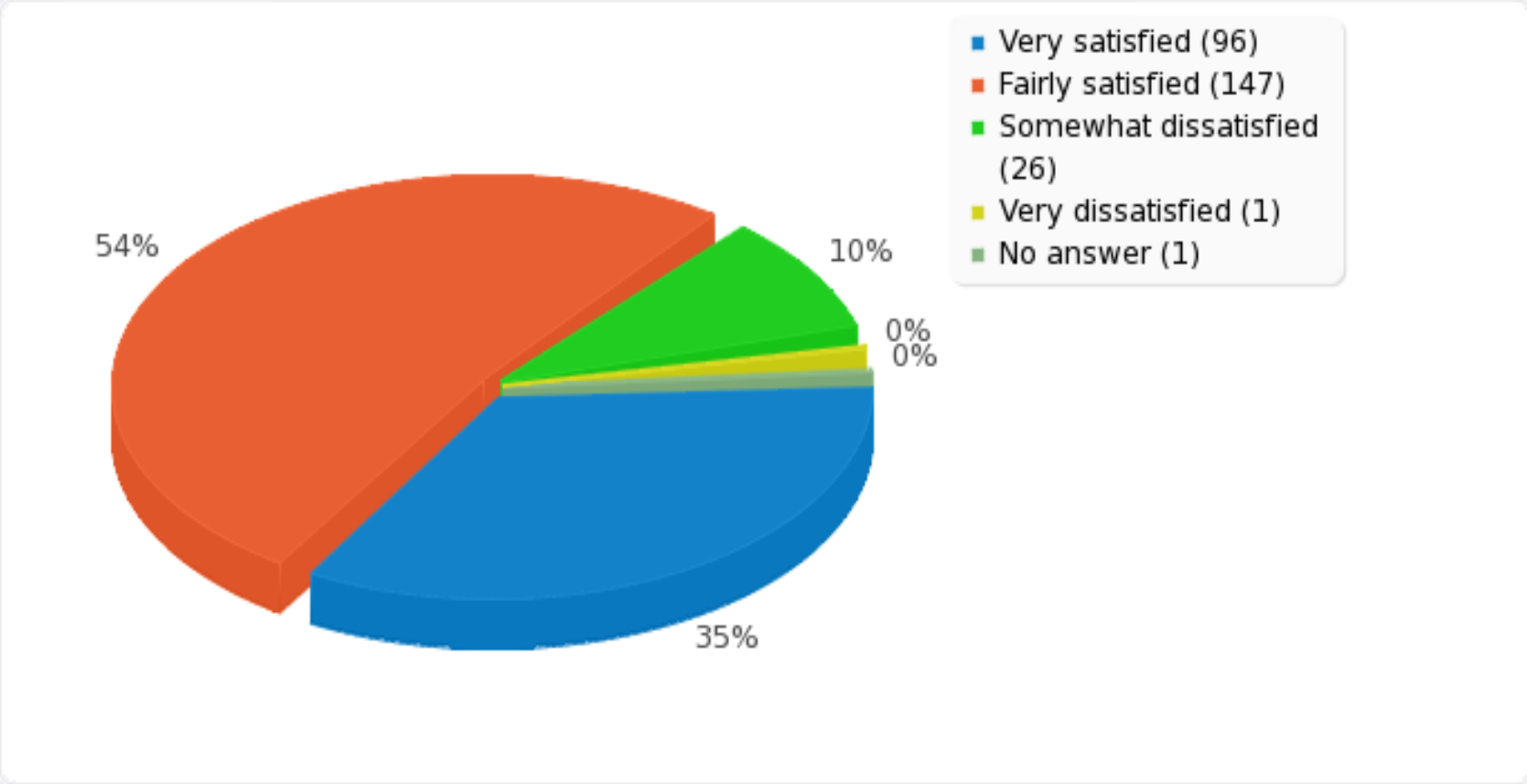
User-friendly utilization of the MicroScope tools?



In case of answers 'Average' or 'Weak', explain why

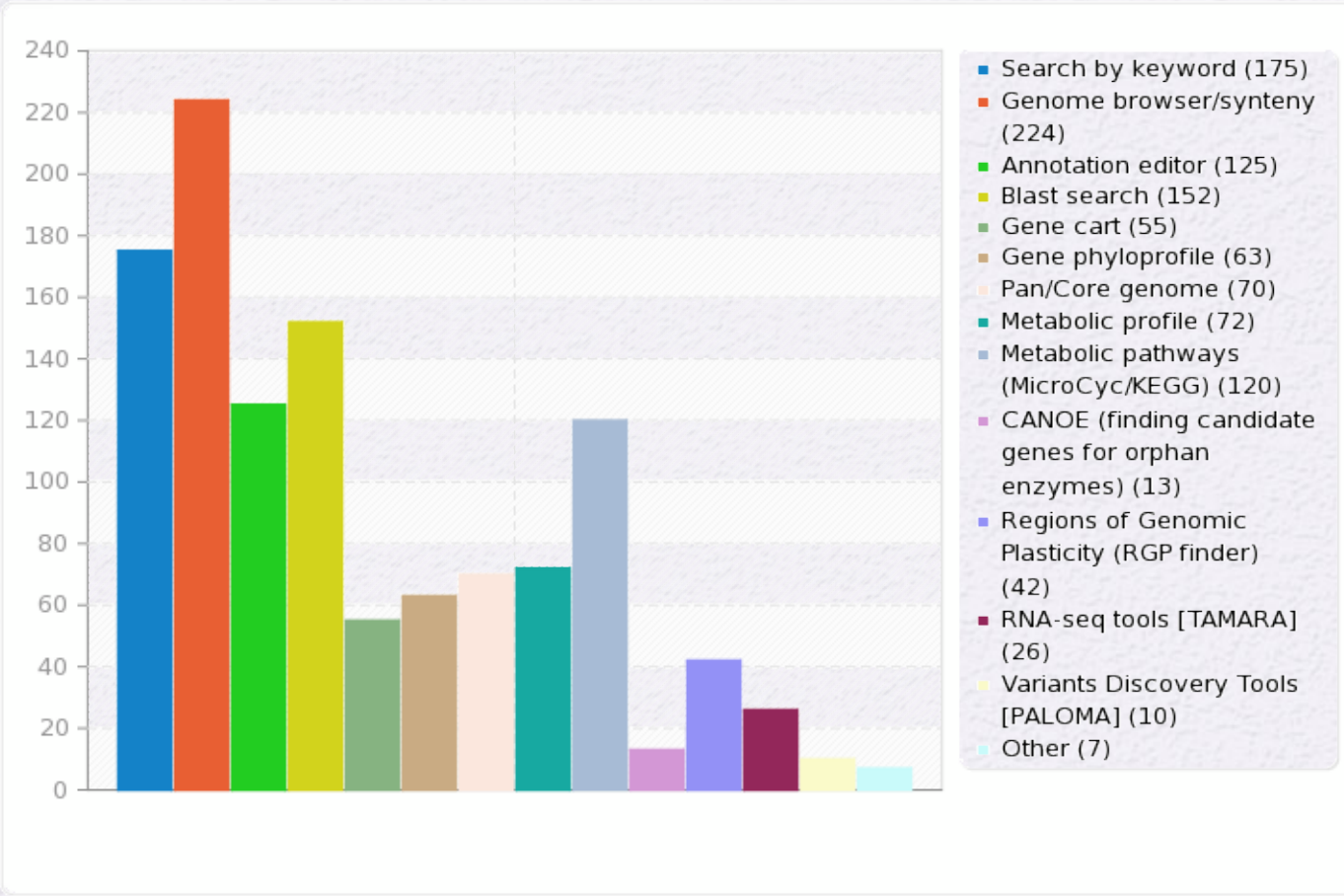
- **clarity of the interface** (5 observations)
  - Not intuitive.
  - sometimes very difficult to find the tools or the data you are looking for.
  - not easy to navigate through the genome browsing + homologs.
- **The documentation section on the various tools needs to be improved**
  - documentation section on the various tools needs to be improved.
  - too little information about the analysis steps and parameters of the pipeline.

Are you satisfied with the speed of the page display?

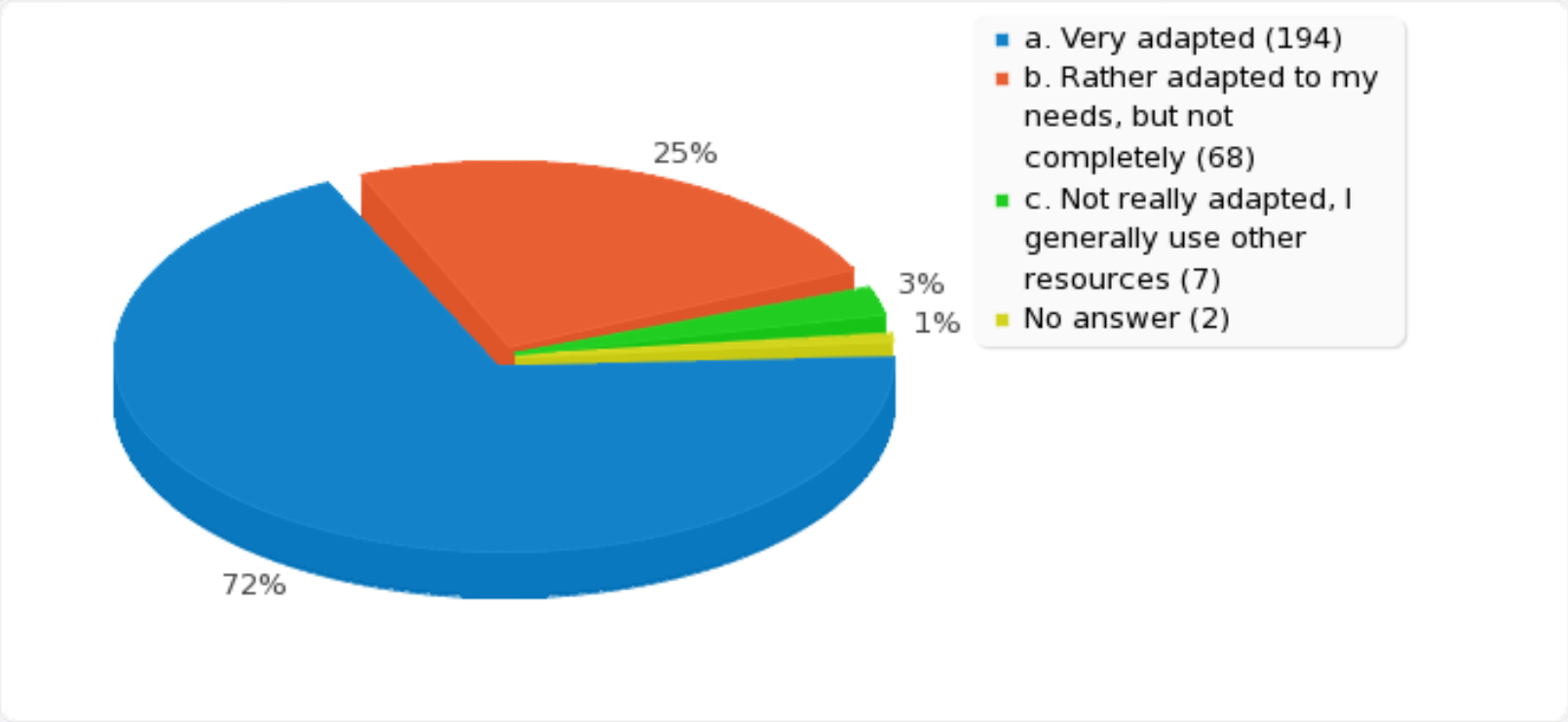


Tools offer

Select the tool(s) you most frequently use



Suitability of our tools for your needs



In case of answer b explain why -1-

- 55 answers.
- **General comments**
  - Some tools are lacking (“but, anyway, I don't think that a comprehensive tool exists nowadays!”).
  - Very good for coding sequence, less for nucleotide analysis
  - Annotation with in house pipelines => faster.

In case of answer b explain why -2-

- **Some tools/option are missing**
  - **Annotation**
    - “Occasionally I use other tools for needs not supplied by MicroScope (e.g. full-scale synteny plots, toxin–antitoxin identification, prophage identification, etc.)”.
    - Easy way to modify start site of genes than Artemis.
    - interpro data linked to gene annotations to show domains and conserved motifs (2 requests).
  - **Comparative genomics**
    - A multiple alignment of fragment shared between two or more genomes.
    - Integration of other tools for orthologs identification (InParanoid, MultiParanoid, …).
    - “it would be nice, to create groups of organism that can then be compared to other groups etc.”

In case of answer b explain why -3-

- **Search /export**
  - homology **searches with HMM profile** (2 requests).
  - Export gene region with flanking sequences of a given length.
  - functions for searching for regulatory motifs / Promotor prediction tools.
  - Download synteny raw data.

- **Assembly**
  - in-built assembly tool (mapping to ref, scaffolding) and gap closure and genome annotation help (2 requests).

In case of answer b explain why -4-

- **Metabolism theme**
  - Translation of metabolic networks to Constraint-Based models (CBM) of metabolism + Integration of tools for stoichiometric analysis of CBM models (FBA, etc).
- **NGS analysis tools**
  - PALOMA-like for variants in favourite/all genomes.
  - Annotation of ncARN; Tamara RNAseq is not directly linked to the DNA Genome browser.
  - Upload operon map and TSS.
  - Direct display transcriptomics data within the genome brower.
  - Better representation of RNA-seq data (heatmap, etc).
  - “all tools for transcriptomics analyses”.

In case of answer b explain why -5-

- **Visualization**
  - A circular genome viewer with genes colored according to different categories (e.g. COG categories,...).
  - Graphical view of predicted signature on protein.
- **Genome brower**
  - **Taxonomic classification.**

In case of answer c, which kind of tools do you also use:

- “Mostly use custom databases, so MicroScope is used in a supporting role (validation, exploration). Tools: NCBI Blast, bioconductor, Chipster, Galaxy”.
- IMG.
- “I use mostly pathway tools and searches of specific genes. I am working on my sequenced genomes and it is complicated to upload my genome in your database”.
- “I spend most of time doing custom analyses”.
- “ Only download PGDBs and do own nucleotide analysis ”.

Lack of knowledge about functionalities

- « Draft genome couldn’t be used in this tools »
- “most tools are available only for referenced genes not for pseudogene due to sequencing errors”.
- import of publicly/private available reference genomes for comparing (4 observations).
- propagate the annotation curation when working on multiple clones of the same species or closely related species (2 observations).
- pathway reconstruction.
- “From a set of strains (20 or more) capacity to obtain the core and the variable genome of each strain as a fasta file and. To obtain from the same set of strains a file for each of them reporting the presence / absence of each gene of the Pan genome.”

Lack of knowledge about functionalities

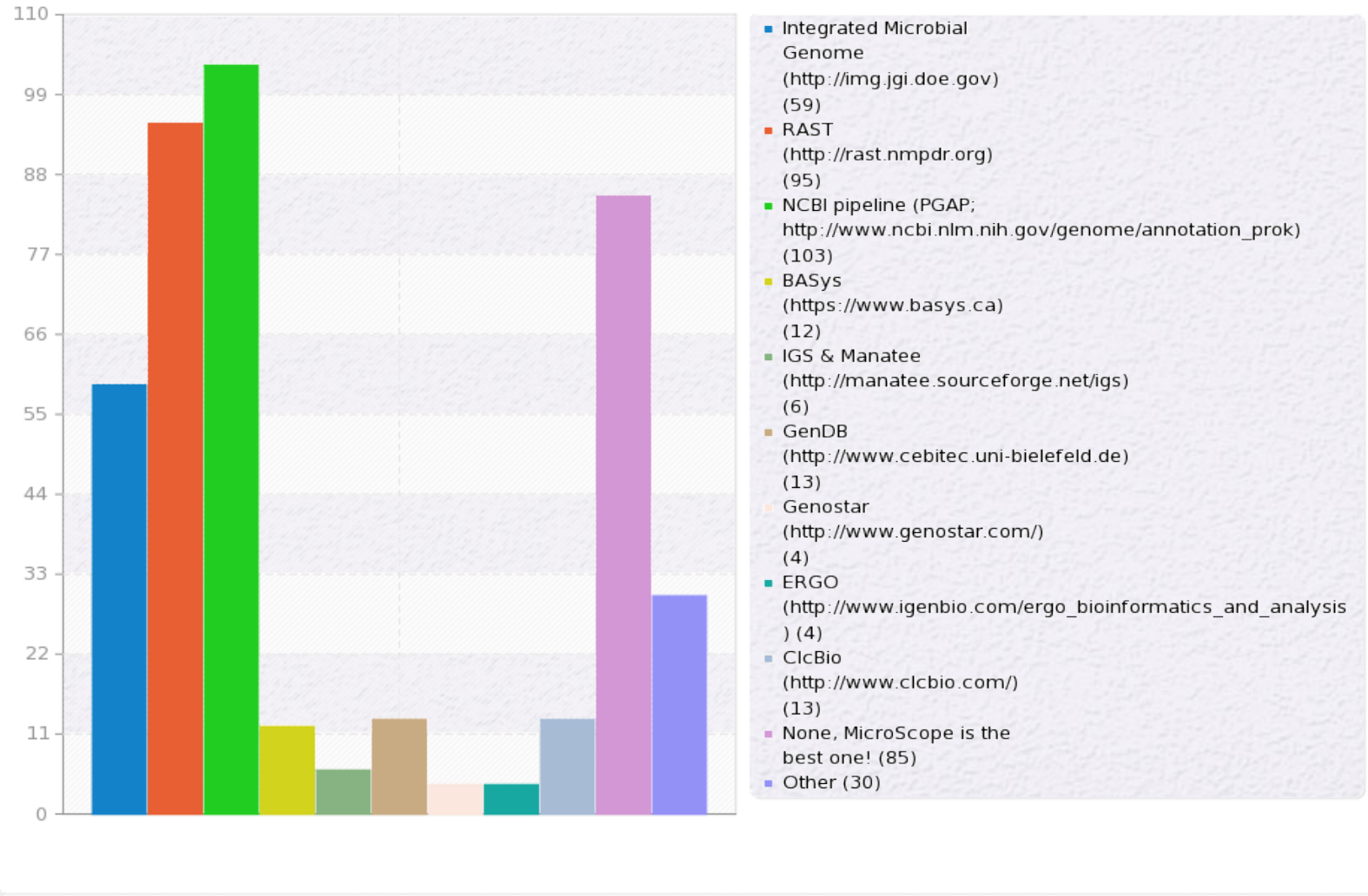
- All the interactive tables should be available as exported tab-delimited tables. Ideally, by a button in the top table row.
- No possibility to browse and/or download data of a very precise region.

About future tools and services

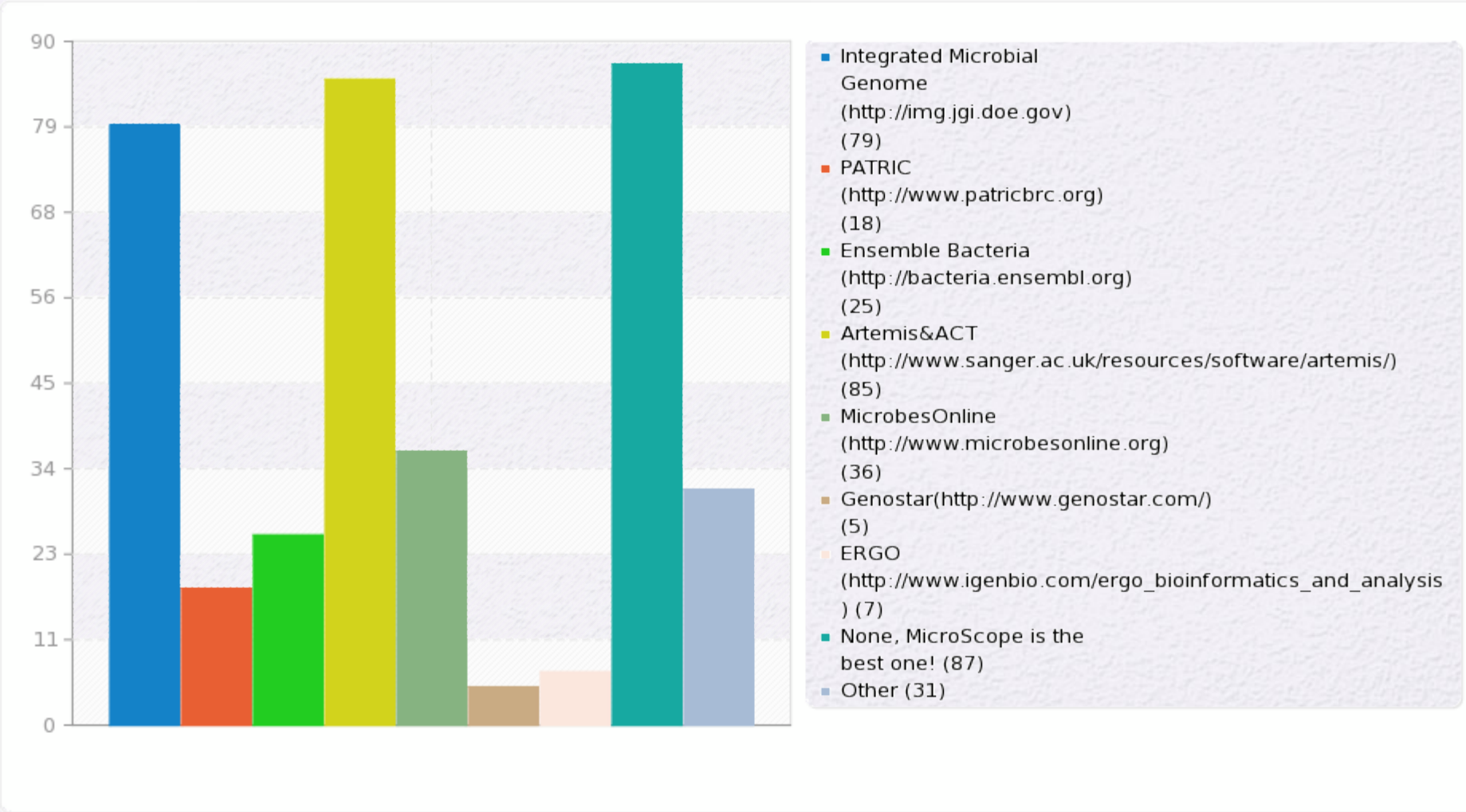
Rate the following tools we plan to integrate into MicroScope :

	I really need it	It will be nice to have it soon	Why not for the next release	No real need	No need at all
Phylogenetic trees from multiple alignments	110(40.59%)	103(38.01%)	37(13.65%)	12(4.43%)	4(1.48%)
Dynamic computation of orthologous clusters of genes from selected bacteria	93(34.32%)	103(38.01%)	47(17.34%)	19(7.01%)	3(1.11%)
Graphical view of predicted domains on a protein	88(32.47%)	91(33.58%)	63(23.25%)	19(7.01%)	3(1.11%)
Flexible circular representation of your genomes	82(30.26%)	95(35.06%)	57(21.03%)	25(9.23%)	8(2.95%)
Run NCBI Blast from the gene Editor	80(29.52%)	98(36.16%)	63(23.25%)	22(8.12%)	3(1.11%)
Genome selection using taxonomic views	64(23.62%)	110(40.59%)	61(22.51%)	28(10.33%)	3(1.11%)
Integration of phenotype data (e.g. Biolog growth phenotype)	60(22.14%)	84(31.00%)	83(30.63%)	32(11.81%)	5(1.85%)
Prophage, CRISPR/cas predictions	52(19.19%)	76(28.04%)	70(25.83%)	43(15.87%)	22(8.12%)
Dynamic SNPs and Ka/Ks computations between pairs of genomes	50(18.45%)	70(25.83%)	56(20.66%)	66(24.35%)	21(7.75%)
API applications for downloading data	49(18.08%)	68(25.09%)	82(30.26%)	47(17.34%)	16(5.90%)

Which other platforms/tools do you use : to automatically annotate your genomes?



Which other platforms/tools do you use : to browse bacterial genomes and to use comparative tools?



We identified 5 main specificities/originalities of the MicroScope platform. How would you qualify them?

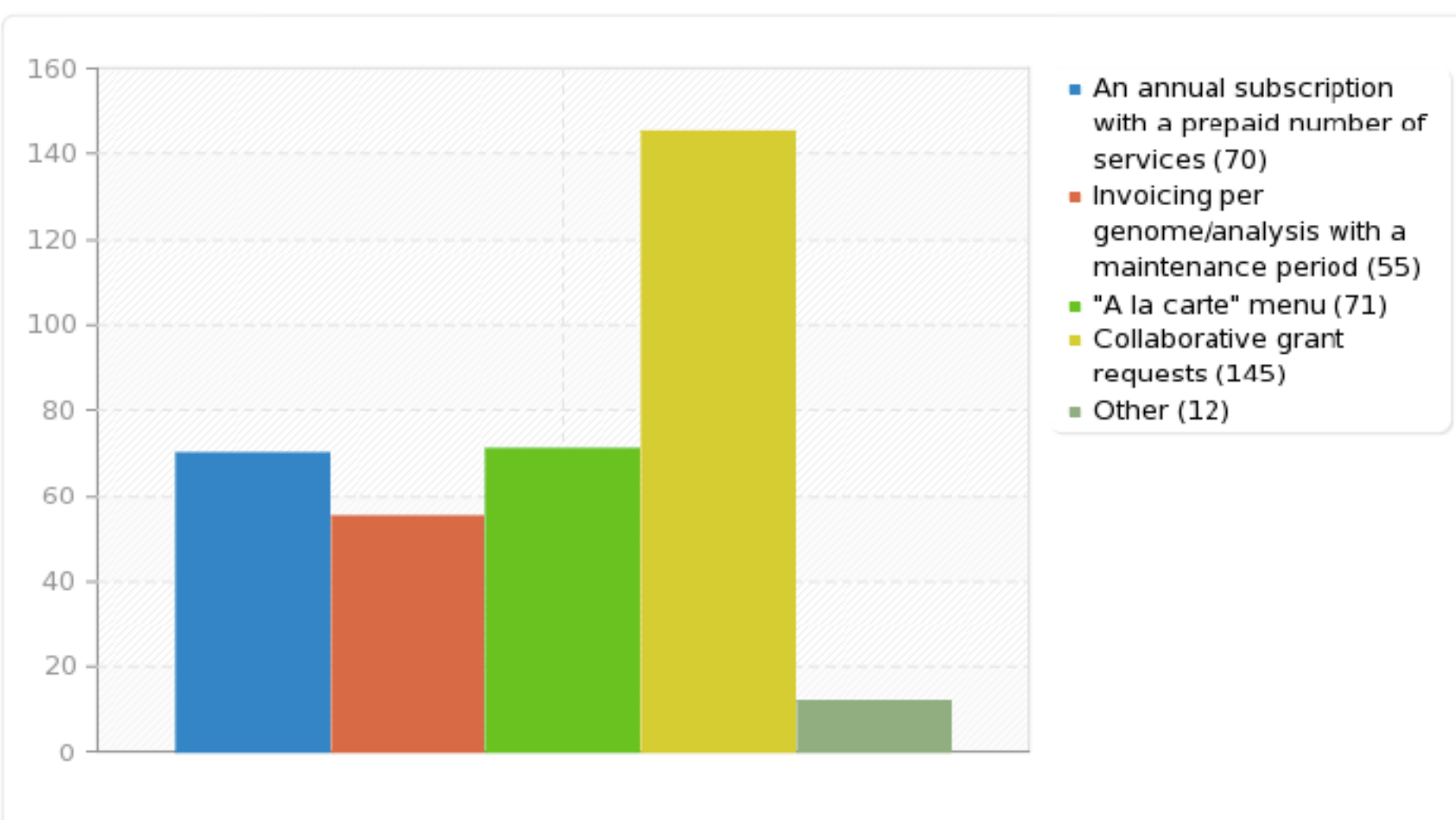
	very important	quite important	important	not really important	not at all important	No answer
Community expert annotations	158(58.30%)	71(26.20%)	32(11.81%)	2(0.74%)	1(0.37%)	7(2.58%)
User Interfaces, especially Synteny visualization	154(56.83%)	74(27.31%)	32(11.81%)	2(0.74%)	1(0.37%)	8(2.95%)
MicroScope team support and expertise	151(55.72%)	79(29.15%)	29(10.70%)	5(1.85%)		07(2.58%)
Tools for metabolism analysis	135(49.82%)	74(27.31%)	47(17.34%)	8(2.95%)		07(2.58%)
Integration of NGS analysis (variant discovery and RNA-seq data)	96(35.42%)	81(29.89%)	66(24.35%)	19(7.01%)	3(1.11%)	6(2.21%)

We plan to develop new services or improve existing ones. How would you consider the following services?

	Essential	Important but not necessary	Useless	No answer
Fast annotation and analysis service	212(78.23%)	45(16.61%)	6(2.21%)	8(2.95%)
Maintenance and updates of computational results	181(66.79%)	78(28.78%)	3(1.11%)	9(3.32%)
Technical and scientific support to research projects	149(54.98%)	109(40.22%)	6(2.21%)	7(2.58%)
Dedicated Web portal for your specific project	107(39.48%)	135(49.82%)	21(7.75%)	8(2.95%)

Mew MicroScope new economic model

Most pertinent way to financially help MicroScope?



Other suggestions

1) government funds.
2) financial involvment of our institutes to garantee the stability over years.
3) dedicated "prestation de service" line in grants.
4) an annual licence fee per site or per user.
5) keep the free platform (4 observations).
6) that is something I have to discuss with my supervisor and whether or not, it is feasible.

How much would you estimate the current cost of the following services?

Fast automatic genome annotation (per genome)]?

Answer	Count	Percentage
< 100€ (A1)	163	60.15%
100€ to 250€ (A2)	69	25.46%
250€ to 500€ (A3)	18	6.64%
> 500€ (A4)	5	1.85%
No answer	16	5.90%

Our estimate : 100€/genome
=> 70 % that's OK ! 13% it's very cheap + 12% too expensive

One year of maintenance with computational updates, up to date reference data and access to expert curation tools

Answer	Count	Percentage
< 100€ (A1)	98	36.16%
100€ to 250€ (A2)	102	37.64%
250€ to 500€ (A3)	41	15.13%
> 500€ (A4)	14	5.17%
No answer	16	5.90%

Our estimate : 200€/genome
=> 59 % that's OK + 30% too expensive !

How much would you estimate the current cost of the following services?

RNA-seq or SNPs variants analysis (per run)

Answer	Count	Percentage
< 100€ (A1)	115	42.44%
100€ to 250€ (A2)	100	36.90%
250€ to 500€ (A3)	34	12.55%
> 500€ (A4)	5	1.85%
No answer	17	6.27%

Our estimate : 150€/genome
=> 60.5% that's OK + 23% too expensive !

"Ad-hoc" developments and specific analyses

Answer	Count	Percentage
< 100€ (A1)	59	21.77%
100€ to 250€ (A2)	67	24.72%
250€ to 500€ (A3)	55	20.30%
> 500€ (A4)	71	26.20%
No answer	19	7.01%

Our estimate : hours cost of an ingenior
=74% that's OK + 16% too expensive

Ideal processing time to get your genome integrated in the platform?

	Few hours	48 hours	1 week	2 weeks	don't care
Extremely satisfactory	**83.03%**	**46.86%**	13.65%	5.17%	3.32%
Rather satifactory	4.43%	33.21%	**32.47%**	12.92%	2.21%
Reasonable	2.21%	8.12%	31.73%	**33.95%**	8.12%
Not really Satisfactory	1.48%	2.58%	9.59%	23.62%	15.13%
Not at all satisfactory	2.58%	2.95%	6.27%	18.08%	**63.84%**

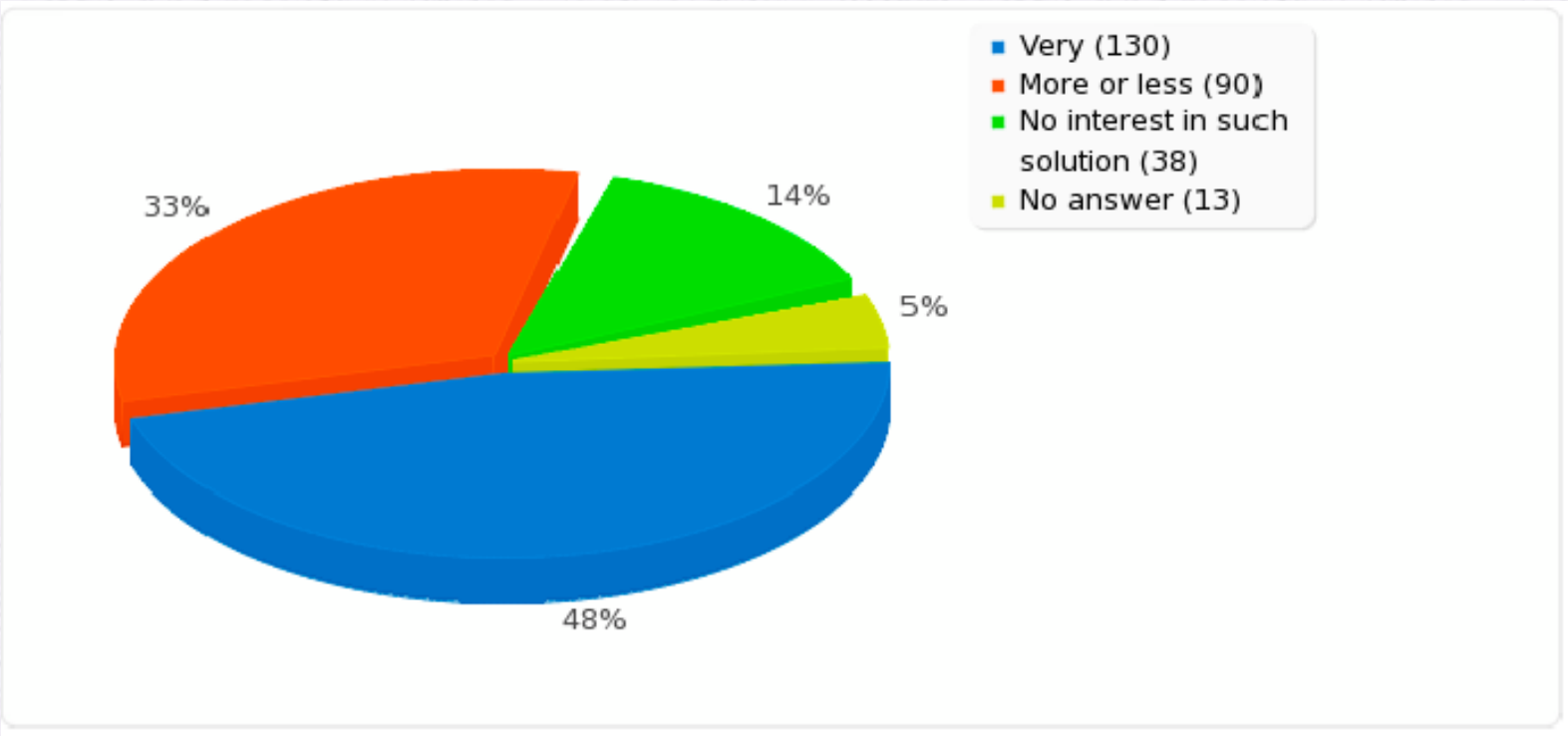
No answer = 6.27%

How long would you wish your genomes to be maintained into the platform?

	1 month	3 months	6 months	1 year	> 1 year
Extremely satisfactory	9.23%	4.80%	3.69%	10.70%	**67.16%**
Rather satifactory	2.95%	5.90%	10.70%	**25.46%**	8.86%
Reasonable	5.17%	11.07%	22.14%	23.25%	8.49%
Not really Satisfactory	13.28%	25.83%	24.72%	18.82%	3.32%
Not at all satisfactory	**63.47%**	**46.49%**	**33.21%**	15.87%	7.01%

No answer = 6%

Would you be interested to install Microscope locally on your own server/computer for an internal use ?



A local installation of MicroScope would be great indeed provided that the installation process would be straightforward enough. A **reasonable annual fee** (lower than the on-line access of course) including software updates could be charged as well.

About local installation

Would the local installation of the service come with the full database (including BLAST hits and synteny hits)? and with the tools to update it with new in-house genomes ?

=> in such a case a **lab-based instance of Microscope could be a very efficient alternative.** It would avoid the enormous and maybe pointless calculation of similarity/synteny searches among bunch of genomes from different projects that people have respective limited interest in.
Also it would likely make the **whole platform much more responsive** (provided the local server is good enough). And then it would releave most of the burden the Genoscope has in terms of database access.
I guess it would **need to have local ingeneers taught how to maintain the database**... a good idea anyway.

Comments on the economic model

1) basic analyses are free, specific analyses and longer maintenance (more than half a year) are premium services.

2) a license fee per site or per user would be much easier for you to manage and less repelling for users. I guess that something around 1000€/yr/site or 200€/yr/user would seem reasonable.

3) What would be the status of our data integrated in MicroScope , as we have not published it yet: do we have to pay money for that too?

4) Our finances are also shrinking, the ANR funding prospects are poor and should the MaGe start asking for large fees, I am not sure we would be able to follow suit.

5) I wish a cooperation-based (ie. including coauthorship on the publication) as is currently available would still exist since not all universities, specially in poorer countries, have budgets which can include the paying of 3rd party services such as bioinformatic analyses.

Comments on the economic model

6) Many of us that works on non-clinical models will not have the funds to include all the genomes that are sequenced (for free most of the times). Given the number of genomes we do, it will be more cost effective to install it locally on our server that to pay annual fees. This is something we have in mind.

7) If fees have to be paid, then the tool must be a real alternative to concurential tools, and I'm not sure people can afford it.

8) I would not subscribe to a fee-based service. I would rather switch to other free annotation platforms e.g. RAST, BASyS, xBase, than have to pay for the service.

9) Our genome data bases have to be maintained for years. If MicroScope should be payed by grants we will be obligated to recover the data and stock and analyse all locally.

Comments on MicroScope

You are happy …

- I'm very convinced on the **key role of Microscope for French researches and worldwide.**
- I think Microscope has **tools which are very easy to use** and which provide **extremely good results** compared to other available tools.. That said, the tools within Microscope have greatly helped me and have made possible the publication of these data in a very good journal !
- Microscope has unabled us to reinforce and initialized international collaborations and annotation consortium. **It is an essential platform for our work.** I would like to thank all the Microscope team.
- **Please stay in business** and - hopefully - not with lots and high costs to individual PIs. Thanks!

You are really happy …

- **Thanks !**
- Overall, MaGe is quite an **excellent platform. Thank you very much.**
- I am a **big fan of MicroScope** - My favourite part of MicroScope is the synteny-based annotation analysis. This is **much better than in other annotation platforms** which are largely homology-based.
- I want to say **"Thank you very much"** to all the members of MicroScope Team for the assistance they are giving to the scientific community all around the world.

But also not happy …

- ORF prediction has become unreliable lately. Some newly integrated genomes contain a lot of overlapping CDS on the opposing strands.
- Outdated BLAST results. Gene prediction tends to overpredict ORFs. Gene carts are largely unusable.
- The big problem with MicroScope is **the turn-over time**. Currently it takes 6-7 weeks to get a genome uploaded and annotated. This is **too slow**. Other platforms take 1-2 days.
- Our highest need is an improvement of **the time it takes to integrate genomes into MaGe.**
- **Limited number of genomes**/ Not updating all genome published (7 observations).
- **The tutorial is not clear enough**!

You got some pb…

- Some tools do not work (the canoe one worked only in ADP1 when I tested it last).
- I had some problems for reaching MicroCyc.
- Phyloprofile blocks when computation too massive.
- import of core genome not always working, seems to be a problem with dealing with large datasets.
- The Blast Search on the microscope website is not anytime working. Also there are often some problems with normal protein format (amino acid sequence).
- Format of data for download not always suitable.
- Gene accession numbers of released genomes do not correspond to the labels in MaGe.
- core/pangenome, I'm not convinced by the results that seems too aggregative and sometimes too exclusive, compared to the BBH results we had that had much more biological meaning.

Others suggestions

- I love the synthenies representation. Would it be possible to have access to a gene information page by left or double clicking on the gene in the synteny on the main genome browser page ?
- An important service is the possibility to export analyses outputs into **figures of high quality for publications**.
- **regular updates** (especially for conserved hypothetical genes) would be precious: at least once a year . It may help to follow global progress in annotation.
- to have possibilities to analyse our own data to compare with the reference strains.
- Improve the **analysis of metagenomes**/ expand to tools to mine and **explore metagenomes** .
- **Programmatic access** like NCBI entrez utilities is necessary, and permanent link as KEGG or BioCyc is recommended.