

# F1000RESEARCH AND DATA PUBLISHING

---

Rebecca Lawrence, PhD  
Managing Director  
24 Feb 2014

[rebecca.lawrence@f1000.com](mailto:rebecca.lawrence@f1000.com)  
<http://f1000research.com>

**@f1000research** | **@rnl\_s**

## SUMMARY

- *F1000Research* introduction
- Data hosting
- Data citation – a question
- Data visualisation
- Data peer review
- Data metrics – a proposal

# F1000 OVERVIEW



## F1000Prime

Find recommended papers



## F1000Posters

Conference poster/slide repository

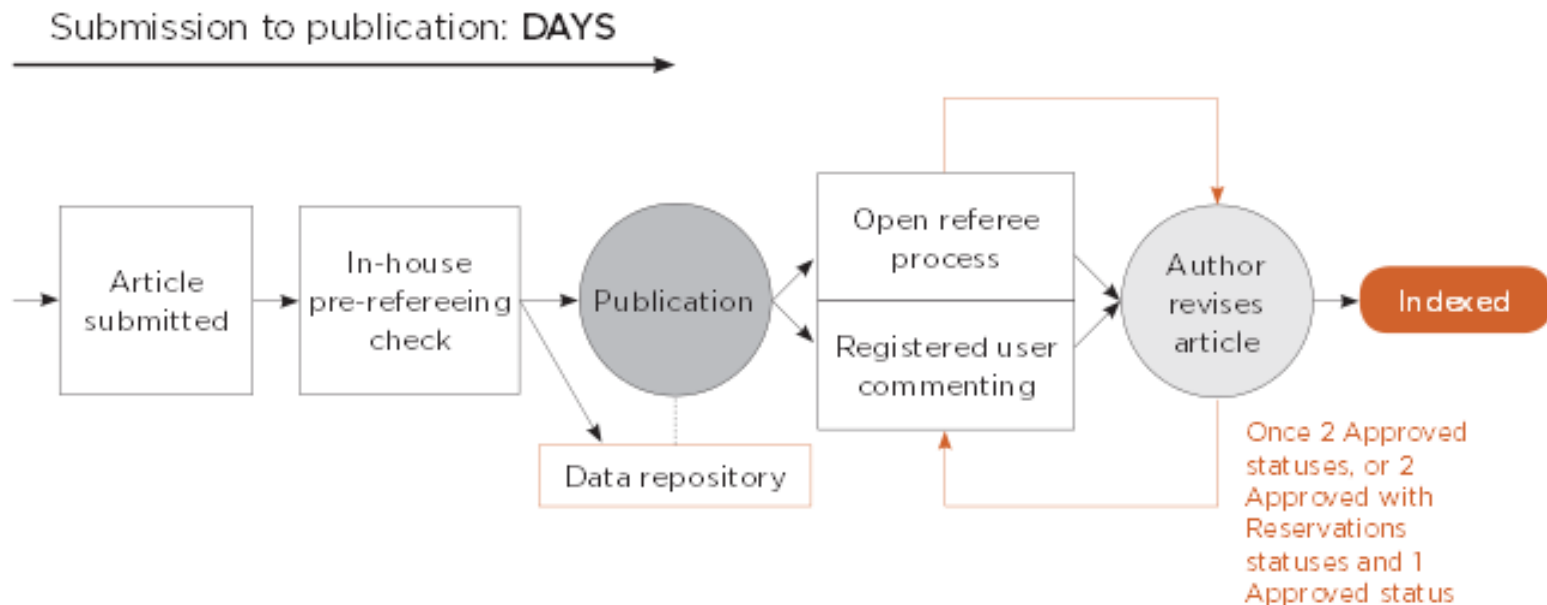


## F1000Research Journal

# F1000RESEARCH: OPEN SCIENCE JOURNAL IN LIFE SCIENCES

Remove the publication delay.  
Invited peer review (post-publication).  
Transparent refereeing.

Inclusion of all data.  
No restriction of access.  
All article types published.



F1000Research » Articles



Version 3 of 3

RESEARCH ARTICLE

REVISED

# Casanovas are liars: behavioral syndromes, sperm competition risk, and the evolution of deceptive male mating behavior in live-bearing fishes [v3; ref status: indexed, <http://f1000r.es/1zi>]

David Bierbach<sup>1,2</sup>, Amber M Makowicz<sup>3</sup>, Ingo Schlupp<sup>3</sup>, Holger Geupel<sup>1</sup>, Bruno Streit<sup>1</sup>, Martin Plath<sup>1</sup>

Author affiliations

Grant information

Views 2713 Download As 395 Cite [social media icons] Share 10 Track

## Abstract

Male reproductive biology can be characterized through competition over mates as well as mate choice. Multiple mating and male mate choice copying, especially in internally fertilizing species, set the stage for increased sperm competition, i.e., sperm of two or more males can compete for fertilization of the female's ova. In the internally fertilizing fish *Poecilia*

## Article Status Summary

### Referee Responses

| Referees                       | 1           | 2           | 3           |
|--------------------------------|-------------|-------------|-------------|
| v1<br>published<br>05 Mar 2013 | ?<br>report | ?<br>report | ?<br>report |
|                                | 1           | 1           | 1           |
| v2<br>published<br>12 Aug 2013 | ✓<br>report | ?<br>report | ✓<br>report |
|                                |             | 1           |             |
| v3<br>published<br>23 Oct 2013 |             | ✓<br>report |             |

1 Katja Heubel, University of Tuebingen, Germany

# DATA HOSTING



Courtesy of Susanna Sansone



- A coherent, curated and searchable registry of repositories, standards, and journal & funder policies in life sciences
- Help stakeholders to make informed decisions:
  - Journals on repositories accredited to the level required by their guidelines
  - Researchers on which journals meet which funder requirements and which repositories meet which journal standards
  - Funders on which journals and repositories meet their policies

## DATA CITATION

We recently added a data and software availability section to all our research articles:

ecological scientists, we envisage that any tools developed will be easily implemented in other research communities, such as the social sciences.

### Data and software availability

#### Data

Figshare: DataUp manuscript data, doi: [10.6084/m9.figshare.884625](https://doi.org/10.6084/m9.figshare.884625)<sup>24</sup>.

#### Software

Zenodo: The DataUp source code package, doi: [10.5281/zenodo.7639](https://doi.org/10.5281/zenodo.7639)<sup>25</sup>.

Bitbucket: Source code for the DataUp Excel add-in and web application, <https://bitbucket.org/dataup/>.

---

Strasser C, Kunze J, Abrams S, Cruse P (2014) **DataUp: A tool to help researchers describe and share tabular data [v1; ref status: awaiting peer review, <http://f1000r.es/2n7>]** *F1000Research* 2014, 3:6

## DATA CITATION - QUESTION

For small data that we host, need a way to cite it:

[Author names] [article year] Dataset [#]. *In*: [article title] *F1000Res* [article volume and number] [dataset DOI]

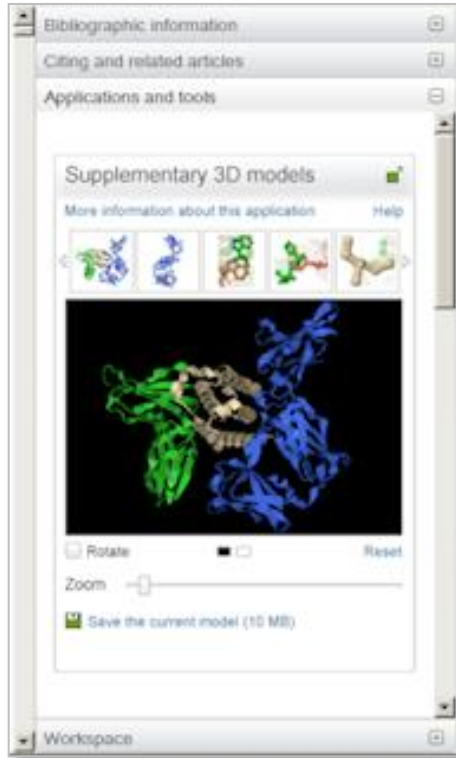
*For example:*

Köhler S, Doelken SC, Ruef BJ *et al.* (2014) **Dataset 1. *In*: Construction and accessibility of a cross-species phenotype ontology along with gene annotations for biomedical research [v2; ref status: indexed]** *F1000Res*, 2:30 (doi: 10.1234/f1000research.1234.d1234)

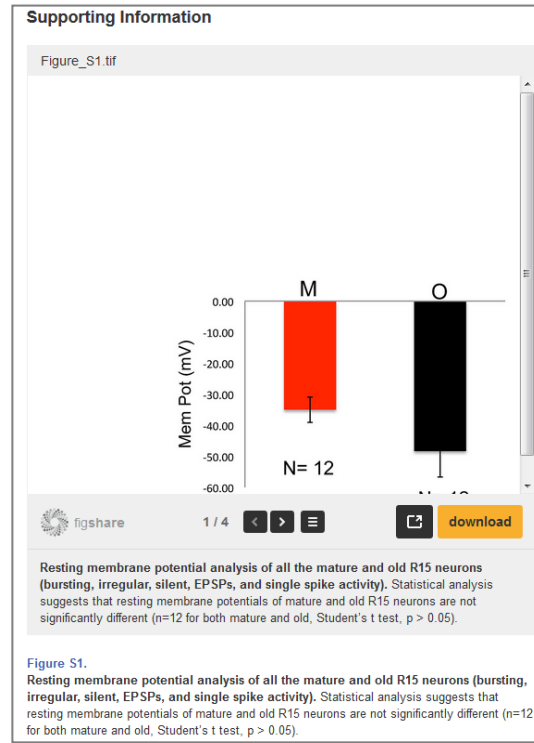
Does this work?



# DATA VISUALISATION: MOVING BEYOND DATA LINKS



**Elsevier**  
PDB and GEO  
links



**PLOS**  
Supplementary  
files

**Son exome files**  
Showing 1/7: Son's Aligned Bam File.bam

| Views | Shares | Downloads |
|-------|--------|-----------|
| 1334  | 5      | 97        |

```

1 FCB021RACXX:4:1208:7911:79502#CAGATCAT 147 1 11941 0 90M = 11883 -147 CTTCCCGTC
2 FCD044UACXX:4:2205:3896:171755#CAGATCAT 99 1 12059 0 90M = 12212 242 ACTGGAGT
3 FCB021RACXX:4:1205:8439:53145#CAGATCAT 99 1 12154 0 90M = 12167 102 ACCACAACC
4 FCD044UACXX:4:1205:1748:199749#CAGATCAT 163 1 12165 0 90M = 12203 127 GCATAGGC
5 FCD044UACXX:4:2103:5744:184901#CAGATCAT 99 1 12167 0 90M = 12274 196 GTAGGGGA
6 FCB021RACXX:4:1205:8439:53145#CAGATCAT 147 1 12167 0 90M = 12154 -102 ATAGGGGA
7 FCD044UACXX:4:1205:1748:199749#CAGATCAT 83 1 12203 0 90M = 12165 -127 TCAACTTC
8 FCD044UACXX:4:2205:3896:171755#CAGATCAT 147 1 12212 0 90M = 12059 -242 CTCACAA
9 FCD044UACXX:4:2103:5744:184901#CAGATCAT 147 1 12274 0 90M = 12167 -196 CCCTCGC
10 FCB021RACXX:4:1107:6633:165696#CAGATCAT 99 1 12275 0 90M = 12313 127 CCTCGCTCC
    
```

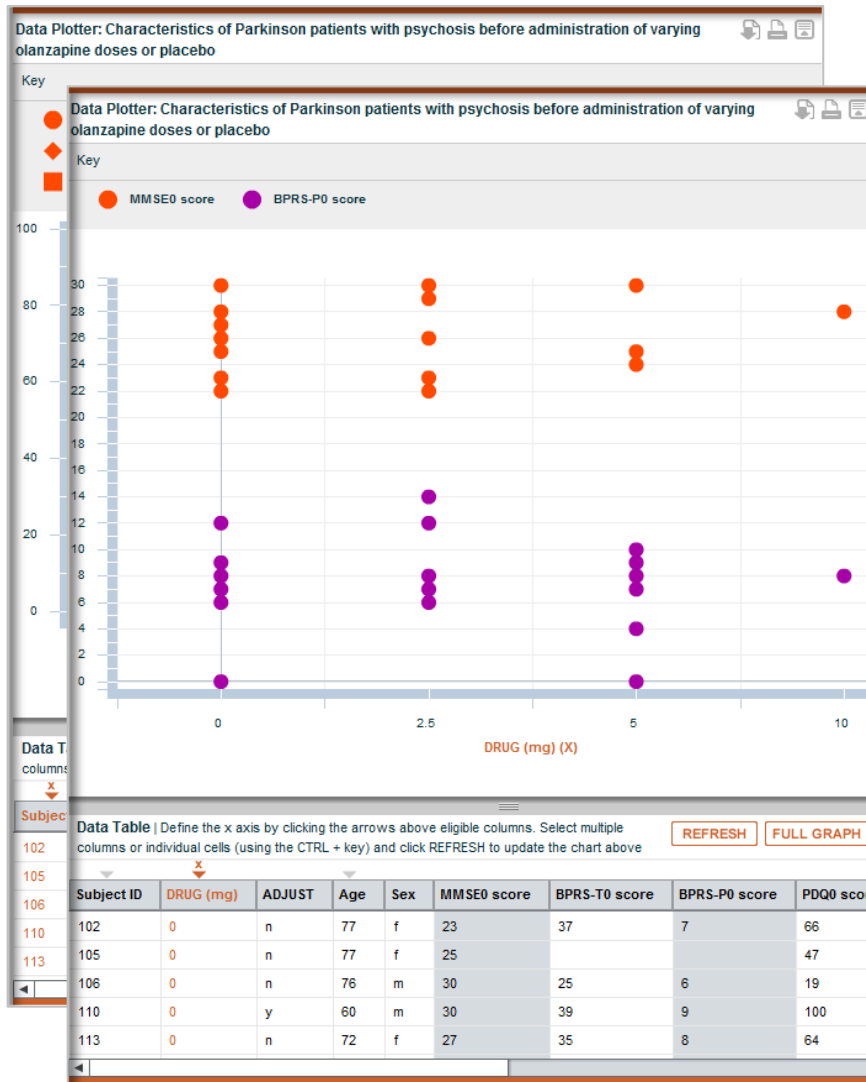
To see the rest of the document click on the icon

figshare 1 / 7

The Fastq files represent the raw exome data for the son. The BAM files are derived from the fastq files by aligning the reads using a Burrows-Wheeler Aligner (BWA). The BAM file (.bam) is the binary version of a tab-delimited text file that contains sequence alignment data. The BAM file index (.bai) provides fast random access to the BAM file. The compressed VCF file (.vcf.gz) describes variant calls of the data in text format.

**F1000Research**  
All data with  
viewers

# IN-ARTICLE DATA MANIPULATION



**A fixed-dose randomized controlled trial of olanzapine for psychosis in Parkinson disease [v1; ref status: indexed, <http://f1000r.es/1au>]**

Michelle J Nichols, Johanna M Hartlein, Meredith GA Eicken, Brad A Racette, Kevin J Black  
*F1000Research* 2013, **2**:150

## *F1000RESEARCH*: DATA REVIEW

Internal pre-publication checks:

- Storage
- Format
- Layout and labelling
- Adequate data?
- Adequate protocol information? (part of NIF trial)

Referees are asked to check:

- Methods were appropriate?
- Adequate information to enable potential replication?
- Format/structure usable?
- Data limitations and sources of error included?
- Does the data 'look' OK?

## DO REFEREES ACTUALLY LOOK AT THE DATA?

“ *In your capacity as referee, did you consider the data as part of your assessment?* ”

| Question   | Percentage |
|--|------------|
| I did not look at the underlying data at all.  | 5%         |
| I looked at the data, but did not consider it when writing my report.                    | 16%        |
| The data formed a part of my editorial decision, but I did not comment on it explicitly. | 50%        |
| I mentioned the data in my referee report.   | 29%        |

## NEED FOR BROADER DATA METRICS

- Need adequate metrics to encourage time to be spent on making data more useable (as opposed to just producing more research).
- Otherwise, large % of funders' money may fund research that no-one else can reproduce or reuse.
- Developing metrics for data articles seems the easiest first step.

## THE PITCH

Identify a set of metrics to enable data output to be measured that:

- Data repositories agree to capture and expose
- Publishers agree to capture and expose
- Funders agree to recognise
- Institutional administration departments agree to recognise
- All agree to make publicly available and share
- Approaches are standardised to enable comparison between sources

And that are significant enough for researchers to be willing to spend adequate time on sharing their data

## WHO WOULD BENEFIT?

### Funders

- ROI on funding towards development of data repository infrastructure
- Capture of a broader set of research outputs from research funding

### Data repositories

- Demonstrate impact of the research being captured to their funders

### Academic institutions

- Capture impact of a broader set of outputs from their researchers
- Increase collaborations

### Researchers

- Priority on their work
- Credit for their data
- Reduce issues of competition between time spent sorting out data versus writing up next paper
- Increase citations from inclusion of data

---

## PROPOSE A WORKING GROUP

- Comprising:
  - Data publishers
  - Institutions with more advanced institutional data repositories
  - Major funders
  - Major data centers/repositories
  - Scientists from data-heavy disciplines
- To:
  - Create a pilot within a specific scientific discipline (life sciences)
  - Agree a set of metrics
  - Agree to implement these metrics across a couple of members of each stakeholder group
  - Agree ways to measure effectiveness of the implementation of the metrics on all the relevant stakeholders
  - Assess success and disseminate as a white paper



---

# Thank you!

[rebecca.lawrence@f1000.com](mailto:rebecca.lawrence@f1000.com)

[@f1000research](#) | [@rnl\\_s](#)