

The Fannie and John Hertz Foundation Application for Graduate Fellowship

To apply for a Hertz Foundation Graduate Fellowship, you must provide us with all of the information requested. Doing so will likely require at least 15 but no more than 30 minutes. Please read the instructions carefully before filling in each box. It is important that you format your answers according to the instructions.

Please fill in your complete name:

Mr.	Tyler	Carter	Shimko	
Mr./Ms./Mrs. First name	Middle name	Last name	Name adjunct	

Enter your date of birth using the format mm/dd/yyyy,

e.g. **June 10, 1983** would be **06/10/1983**:

Date of birth

Enter the name of the country where you were born:

Country of birth United States

If you are you a US citizen, print US in the Citizenship box.

If you are a Permanent Resident, print PR.

US

Citizenship?

If you are neither a US citizen nor a Permanent Resident of the US, you are not eligible to apply for a Hertz Fellowship. (If you are not a US citizen, we will ask to see proof of Application for Citizenship or your Permanent Resident Card when you are interviewed).

Current School/Work Information

If you are currently attending school, enter that school's name in the "Present school" box. If you are currently out of school, put your current occupation in the "Currently out of school" box and leave the "Present school" box blank. In the City and State boxes, enter the city and state of your current location.

Univ Utah

Present school

Currently out of school

Please enter the location, (city and state only), of your present school in the next box.

Salt Lake City

UT

Present school location (city, state)

What is your current status? Undergraduate

Current Status

Undergraduate Institution

If you are currently a Graduate student, name the school that awarded your undergraduate degree. If you are currently an Undergraduate, enter your current school.

Univ Utah
UG School

UG School (if not listed)

Fall 2015 - Spring 2016 will be your first year as a graduate student.

What is the Academic Major (field of study) in which your undergraduate degree was awarded, or, if you are an undergraduate, the field in which you are currently enrolled? Academic Major means: physics, chemistry, mathematics, mechanical engineering, etc.

Biology

Undergraduate Academic Major

Overall Undergraduate GPA (4.0 = A)

3.89

UG GPA

3.75

UG GPA Major

What is your undergraduate GPA in your major subject?

If your undergraduate university does not assign grades, please enter "not available." Your scores in the GRE Aptitude and Advanced Subject Tests will be especially important to the foundation in evaluating your application.

If you are currently a graduate student, what is your overall graduate GPA?

Grad GPA

Graduate Institution

If you know the educational institution at which you currently propose to hold your Fellowship, please name it. Otherwise select, "Not yet known." If the school you select is not currently included in our list of Currently Participating Institutions (page 6 of the Instructions), please understand the process by which that school will be approved.

It is not necessary for you to have been accepted at this institution at the time you submit your application. While a definite school choice will be required at the time of Fellowship acceptance, this choice may subsequently be changed by submitting a written request to the Foundation.

Not yet known
School Choice

School Choice (if not tenable)

In the next box, please indicate what your proposed General Field of Study will be at the school selected above. Please refer to the instructions for a discussion of Hertz-tenable fields of study.

Quantitative Biology
General Field

Please note that the Foundation does not support students seeking professional degrees other than the PhD, (e.g., MD, LLB, MBA, etc. are not supported).

In the Specialty Area box, please indicate what specialty area you intend to study, (e.g. condensed matter physics, molecular biology-DNA analysis, electrical engineering). If you do not know what Specialty Area you'll be studying, enter "Unknown."

Quantitative Genetics
Specialty

Graduate Institution (continued)

Quantitative Biology

General field

Quantitative Genetics

Specialty area

Please list other graduate schools that you are seriously considering attending (in addition to your choice listed above):

Stanford University
University of California, San Francisco
University of California, San Diego
University of California, Los Angeles
University of Washington

Other Potential School Choices

Toward what **ultimate** degree will you be working while in Graduate School?

PhD

Degree sought

During what calendar year do you expect to complete your Ph.D. studies?

2020

Completion date

Please list other multi-year Fellowships you have applied for, currently hold, or are planning to apply for:

NSF Graduate Research Fellowships Program (Applying)
DOE Computational Science Graduate Fellowship (Applying)

Other Multi-Year Fellowships

Preparation for Graduate Work in the Applied Physical Sciences

You should have had university-level courses in physics, mathematics and chemistry beyond the introductory level. Please indicate below the number of upper division courses in each category that you have taken already or plan to take as an undergraduate. Enter the numbers as decimals, with a year-long course counting as 1.0, a semester course as 0.5, and a quarter course as 0.33. Please count research courses as upper division courses.

Upper-Division Course Years

Physics

0

Chemistry

1

Engineering
(ME, ChE, EE, CE)

0

Mathematics

1

Computer Science

0

Earth Science

0

Modern Biology

5

Your Location

During the application process, it is important for us to be able to reach you on short notice. We may need to arrange a technical interview, or we may have a question about your application. Please give careful thought to this information. Further along on this form you will have the opportunity to give several alternate telephone numbers and addresses where you might be reached during the next 6 months. If any of this information changes between the time you submit this application and **March 15, 2015**, please inform us promptly of the changes, by e-mail to askhertz@hertzfoundation.org.

What is your present **mailing** address?

Present street address

City	State (abbr only)	Zip (first five digits only)
------	----------------------	---------------------------------

In which state do you currently hold residency?

Residency

What are your present home and work phone numbers? (Be sure to include area codes.)

Cell phones

Home phones

Work phones

e-mail address

Do you have a second or alternate e-mail address? If so, enter it here:

Alternate e-mail address

We have found e-mail to be an especially good way to reach Fellowship applicants. We will be using e-mail to correspond with you regarding the status of your application, and to let you know when your application arrives at our office. We may also be using e-mail to alert you to the possibility of a technical interview.

Until what date do you expect this contact information to be valid? Use the same format, **mm/dd/yyyy**, as was used to enter your day of birth (e.g. June 10, 2007 would be **06/10/2007**).

Contact information
valid until:

In order to plan the technical interviewing process, we need to know what airport is nearest to your current location. Find and enter the most appropriate airport abbreviation (from the list included in the Instruction section) into the "closest airport" field.

Closest airport

Please enter the address where we should mail the final selection announcement letter on/before April 1, 2015.

April 2007 street address

City	State (abbr only)	Zip (first five digits only)
------	----------------------	---------------------------------

The Interview

You will be contacted in mid-November to schedule a technical interview if you are one of the approximately 20% percent of our applicants who are selected for continued consideration. This interview generally lasts 45 to 60 minutes. It is patterned after the PhD oral exam and you may be asked to perform calculations, discuss your previous research work, and to demonstrate the breadth and depth of your technical knowledge.

The interview may be held in one of a variety of locations, including hotel conference rooms, restaurants, airport conference rooms, or on a university campus. You may not be given much advance notice and you will have to arrange transportation to your interview. Your reasonable-and-ordinary travel expenses will be fully reimbursed by the Foundation if, at our request, you must travel over 5 miles. If you are in any of the following cities or broad surrounding regions, interviews will be held on the following dates in the cities indicated. You will be notified of the actual location when you are contacted for an interview.

Saturday, November 22, 2014, Boston

Saturday, December 6, 2014, Chicago

Saturday, January 17, 2015, Philadelphia

Saturday, January 17, 2015, San Francisco Bay Area

Saturday, January 24, 2015, Los Angeles

We expect to interview most applicants at one of these five venues, but interviews at other locations may sometimes be scheduled by the Foundation. Reasonable expenses will be paid for your travel to the nearest interview site. While you may choose to travel to a more distant site at your own expense, you may wish to keep these dates in mind when making your Fall/Winter travel plans.*

Approximately 50 Finalists will be selected in early February and these Finalists will be interviewed for a second time between early February and mid-March. The finalist interviews will take place as follows:

February 14-16, 2015 in the San Francisco Bay Area; or

February 27-March 1, 2015 in Boston

If you are selected as a Finalist, you will be asked to travel to the venue closest to your proximity. You may not be given much advance notice and you will need to arrange transportation to your interview. Your reasonable-and-ordinary travel expenses will be fully reimbursed by the Foundation if, at our request, you must travel over 5 miles.*

* We ask that you consult with us first if your travel expenses are likely to exceed \$100.

Four very brief essays

In the spaces below, (i.e., in approximately 300 words), please provide concise responses to each of the following questions.

Question 1: Choice of Field and Future Expectations

How did you choose your field and what are your primary expectations of your future career? If you are currently in your second or later year of graduate school, you should make your case here for why receiving a Fellowship would result in exceptional leverage in the kind and quality of your graduate work, including your ability to pursue promising new ideas. Please understand before continuing this application, that such a case will have to be very strong to be considered further, and that new Hertz Fellowships are very rarely granted to students currently in their second year of graduate study or beyond.

Quantitative genetics, the study of traits that vary continuously such as height, connects the physical traits of an organism to its genetic material. Recent advancements in DNA sequencing and phenotype data collection have greatly refined the field's methods and scope. Though quantitative genetics has existed for nearly a century, it is only with these new sources of data that we can begin to apply its methods to issues of human health and society.

I have chosen to work in quantitative genetics because it has the ability to revolutionize many facets of human life, including agriculture, health, and environmental wellness. Quantitative genetics allows us to explore the interplay between the genetic material of an organism, be it a plant, fungus, or animal, and the organism's environment. I have three main goals for my career in quantitative genetics.

My first goal is to participate heavily in collaborative and interdisciplinary academic research. Through collaborations with experts in biochemistry, molecular, and systems biology, I will learn protein dynamics and biological network analysis to predict the molecular mechanisms of variation and develop comprehensive models of genetically complex traits, incorporating both genetic and environmental factors.

My second goal is to instruct the next generation of scientists and citizens. My passion and skills in biology came from patient and committed mentors, and I want to give that same attention to others. I will teach young biologists, both in the classroom and the laboratory, about the most advanced techniques in the field. I will also translate my research to be comprehensible by the general public. By fostering a greater understanding of science and society, the efforts of scientists will likely see increased political and financial support.

Question 2: Proposed Field of Study

How do your proposed field of study and career constitute an application of the **physical** sciences or engineering?

Quantitative genetics utilizes mathematics and computer tools to analyze the underlying genetic networks of complex traits, such as human height and disease susceptibility. However, to fully comprehend the interconnected web of molecular interactions that govern phenotypes, methods of chemistry and physics must also be employed. Gene expression is commonly induced through the binding of genetic regulatory elements nearby the sequence encoding a gene. Additionally, genes often have multiple alleles, involving different changes in amino acid sequence that each confer a distinct effect on protein structure. Both of these critical DNA sequences affect an organism's phenotype and can be identified as quantitative trait loci (QTL) using accepted genome wide association study (GWAS) techniques.

Although GWAS techniques can identify causative mutations, we must utilize principles of thermodynamics and biochemistry to determine how changes in DNA or amino acid sequence affect regulatory element binding or protein folding, respectively. In order to discern the specificity of regulatory elements, the most thermodynamically stable binding configuration must be elucidated. Moreover, similar thermodynamic stability analyses must be performed in order to determine the role of mutations in changes to protein structure.

Throughout my career, I plan to link the field of quantitative genetics with that of structural biology. In order to accomplish this feat, I will need to understand and apply principles of the physical sciences such as thermodynamics. By approaching the problems facing quantitative genetics from an engineering perspective, recognizing output phenotype as the result of a series of complex and interconnected molecular interactions, we will be able to draw even greater predictive power from GWAS techniques. This goal will require new, more efficient computational tools in the fields of quantitative genetics and structural biology. I wish to combine my knowledge of the physical and chemical sciences with that of biology and computer science to construct these tools.

SUBMIT THIS
APPLICATION
ONLINE

Question 3: Choice of Graduate School

What are the considerations involved in your choice of graduate school?

There are three major considerations that will collectively dictate my choice of graduate school. The first of these aspects is a strong interdisciplinary culture that will allow me to learn from the best geneticists, statisticians, biophysicists, and evolutionary biologists. In order to create new knowledge from findings in quantitative genetics, I must know how to design experiments, properly analyze resulting data, and predict alterations to molecular structures. I have carefully selected candidate schools which house leading researchers in all of the above fields and encourage interdepartmental collaboration to take ideas from theory to application.

My second major consideration is the availability of the computational resources necessary to conduct novel research in quantitative genetics. Modern quantitative genetics relies on the most advanced computational equipment to perform incredible numbers of calculations in a timely fashion. By attending a school with the appropriate computer hardware and instructional resources, I will prime myself to make fundamental contributions to the field. All of the schools that I am considering have excellent computer science or computational biology departments with bleeding-edge hardware and unparalleled expertise.

My final consideration is location. While not a direct measure of the potential of a school to provide a stellar education, schools close in proximity to top medical facilities, scientific and technological hubs, or other academic institutions multiply their capacity for collaboration. Since my interests extend beyond the academic study of quantitative genetics, it will be critical for me to form connections outside of the traditional scientific realm. I am considering schools located within or adjacent to major urban scientific hubs so that I will have the opportunity to collaborate with and learn from individuals with expertise in a wide array of pertinent disciplines.

SUBMIT THIS
APPLICATION
ONLINE

Question 4: Chronological Resume

Provide a concise resume, in chronological order, with dates, recapitulating significant periods of technical and other creative activity since high school graduation. Omit activities only distantly related to your professional development. Include workshops, summer schools, a general description of all courses of study pursued (e.g. "3 quarters of Differential Equations") and degrees expected or awarded (dates, institutions, fields). Separate your undergraduate activities from your graduate activities (if/as applicable) with a single dashed line.

Honors Bachelor of Science in Biology from the University of Utah (Expected May 2015)

Course Highlights:

1 semester of Object-Oriented Programming
4 semesters of computer labs in the R language
1 semester of a computer lab in the Python language
2 semesters each of Probability and Statistics, Calculus
2 semesters each of Organic Chemistry, Biochemistry
2 semesters of Physics for Scientists and Engineers
1 semester each of Genome Biology, Molecular Biology, Human Evolutionary Genetics

Research Experience:

08/2014 to Present - Undergraduate Researcher (Stanfield lab, University of Utah, 20-25 hrs/week) - Analyzed genomic sequence data and identified genetic variants implicated in abnormal sperm activation phenotypes in *C. elegans*. To be presented at the University of Utah's 2015 Undergraduate Research Symposium and published as my Honors Thesis.

05/2014 to 08/2014 - Undergraduate Researcher (Andersen lab, Northwestern University, 40-50 hrs/week) - Designed and built software to process data from COPAS large-particle flow cytometers and investigated the genetics of complex traits through the analysis of genetic linkage mapping data from previous summer. Published in Shimko and Andersen, 2014 (PLOS ONE) and presented to the Northwestern University Worm Club on July 30th, 2014.

08/2013 to 05/2014 - Undergraduate Researcher (Stanfield lab, University of Utah, 20-25 hrs/week) - Genetically mapped mutations affecting sperm activation in *C. elegans*.

05/2013 to 08/2013 - Undergraduate Researcher (Andersen lab, Northwestern University, 40-50 hrs/week) - Constructed high-throughput phenotyping pipeline for use in genetic linkage mapping studies with *C. elegans*. To be published in a paper that is currently in preparation.

01/2013 to 05/2013 - Undergraduate Research Advisor (Office of Undergraduate Research, University of Utah, 5-10 hrs/week) - Helped fellow undergraduates find research opportunities on campus and begin involvement in research projects.

08/2012 to 05/2013 - Undergraduate Researcher (Jorgensen lab, University of Utah, 10-20 hrs/week) - Investigated protein involved in synaptic vesicle recycling in *C. elegans*.

05/2012 to 08/2012 - Undergraduate Researcher (Kruglyak lab, Princeton University, 40-50 hrs/week) - Constructed *C. elegans* strains used to investigate the roles of genomic regions on phenotypes in response to the herbicide paraquat. Presented by Dr. Erik Andersen at the 2013 International *C. elegans* Meeting and the 2013 Society for Molecular Biology and Evolution Meeting. Presented by myself at the 2013 Utah Conference on Undergraduate Research and the 2013 National Conference on Undergraduate Research. To be published in a paper that is currently in preparation.

Academic Honors

In the space provided below, list, in chronological order, academic honors and distinctions you have received and the time or time-interval of receipt. Separate your undergraduate from your graduate awards (if/as applicable) with a single dashed line (Include title, reason for award, and where/when received). Especially significant awards received in high school can also be included. Use no more than one line per award whenever possible (what, where/when received).

Expected upon graduation (May 2015) – Undergraduate Research Scholar Designation –
Recognizes 2 semesters (minimum) of research with presentation and publication

10/2014 – Marshall Scholarship Candidate – Officially endorsed by the university to
compete for the Marshall Scholarship

10/2014 – Churchill Scholarship Candidate – Officially endorsed by the university to
compete for the Churchill Scholarship

09/2012 to Present – Undergraduate Research Ambassador – Selected to present on research
at University functions

Fall 2011 to Spring 2014 – Dean's List – Minimum 3.5 GPA each semester

**SUBMIT THIS
APPLICATION
ONLINE**

Fellowships, Scholarships, etc.

List here, in chronological order, any fellowships, scholarships, teaching or other appointments held since entering college or university. Separate your undergraduate and graduate education intervals (if/as applicable) with a single dashed line. Put an asterisk (*) at the beginning of all lines indicating a national-level award (e.g., award of a National Merit Scholarship; NSF, NASA, or NDSEG Fellowship; election to Phi Beta Kappa, Tau Beta Pi, or Sigma Xi; etc.). Use one line per item whenever feasible (provide basic what, where, when data).

Fall 2014 to Spring 2015 – Myriad Genetics Academic Excellence Award (Competitive, Merit-Based, Offered through University of Utah College of Science)

* Fall 2013 to Spring 2015 – Barry Goldwater Scholarship (Nationally Competitive, Research-Based, Awarded as a sophomore)

Fall 2013 to Spring 2014 – University of Utah College of Science Dean's Scholarship (Competitive, Merit-Based)

Fall 2013 to Spring 2014 – Theodore Verender Hanks Scholarship (Competitive, Merit-Based, Offered through University of Utah College of Science)

Fall 2012 to Spring 2015 – Full Resident/Half Non-Resident Partial Tuition Waiver Scholarship (Merit-Based, University-Wide Award)

Spring 2012 – Undergraduate Research Opportunities Program Assistantship (Competitive, Research-Based, University-Wide Award)

Fall 2011 to Spring 2012 – Full Resident Partial Tuition Waiver Scholarship (Merit-Based, University-Wide Award)

SUBMIT THIS
APPLICATION
ONLINE

Previous Research

Please list the most significant research projects that you have pursued, in chronological order. (Include reference information for those that have been formally documented, presented at a conference, or submitted for publication.)

Examples:

- 1) "Controlled Fusion in the Shadow of the French Alps: My Summer Internship at Grenoble"; June-August 2002
- 2) "Peculiarities in Gene Transcription Regulation in wingless Drosophila Mutants", prepared/presented as part of the Student Research Seminar Series; U of Calif. Report, UCRL Report 1029-04, Sept 01-May 02.
- 3) "Deficiencies in Stereospecific Iodination of Thyroxine Precursors Isolated From Chernobyl-Region Voles", J. Exotic Endocrine Chem. 239, 3365 (June 2003), K. Early, M. Y. Name, L. Late, and M. Middle.

*****IMPORTANT*****

Finally, choose one or two projects that best exemplify your own creativity and discuss in more detail what you personally contributed to them.

Note that compelling example(s) of personal creativity are a very important factor in our selection process. Please highlight what makes your personal contributions stand out, and use the first person wherever appropriate. Specific evidence of your personal creativity is more important to us than what your research group did as a whole, unless you contributed centrally to its leadership.

Please submit copies of your most significant scientific publications/reports using the upload capability on Page 13.

Number of:

B.S. Researches Pursued

6

B.S. Researches Documented

5

B.S. Researches Submitted to Refereed Publications

1

Grad. Researches Pursued

Grad. Researches Documented

Grad. Researches Submitted to Refereed Publications

Previous Research

Research Experiences:

- 1) 08/2014 to Present - PI: Gillian Stanfield, University of Utah - "Mapping suppressors of the *swm-1* mutant sperm activation phenotype in *C. elegans*"
- 2) 05/2014 to 08/2014 - PI: Erik Andersen, Northwestern University - "Elucidating genes controlling chemical response in *C. elegans* using recombinant inbred advanced intercross lines"
- 3) 09/2013 to 05/2014 - PI: Gillian Stanfield, University of Utah - "Mapping suppressors of the *swm-1* mutant sperm activation phenotype in *C. elegans*"
- 4) 05/2013 to 08/2013 - PI: Erik Andersen, Northwestern University - "Using high-throughput fitness assays to decipher the genetic causes of *C. elegans* drug sensitivities"
- 5) 09/2012 to 05/2013 - PI: Erik Jorgensen, University of Utah - "Mapping suppressors of the *unc-41* mislocalization phenotype in *C. elegans*"
- 6) 05/2012 to 08/2012 - PI: Leonid Kruglyak lab, Princeton University - "Identifying the genes that control paraquat resistance in the roundworm *C. elegans*"
- 7) 09/2011 to 05/2012 - PI: Erik Jorgensen, University of Utah - "Universal Transgene Insertion in *C. elegans*"

Publications (* denotes peer-reviewed):

- 1) * Tyler C. Shimko and Erik C. Andersen. (2014) "COPASutils: An R Package for Reading, Processing, and Visualizing Data from COPAS Large-Particle Flow Cytometers." PLOS ONE. DOI: 10.1371/journal.pone.0111090
- 2) Tyler C. Shimko and Erik M. Jorgensen. (2012) "Universal Transgene Insertion in *C. elegans*." University of Utah Undergraduate Research Abstracts Journal, Volume 12.

Presentations (* denotes presenter):

- 1) 7/2014 - Tyler C. Shimko*. "Linkage mapping with recombinant inbred lines." Northwestern University Worm Club.
- 2) 7/2013 - Tyler C. Shimko, Robyn E. Tanny, and Erik C. Andersen*. "Using high-throughput fitness assays to decipher the genetic causes of *C. elegans* drug sensitivities." Society for Molecular Biology and Evolution Meeting.
- 3) 6/2013 - Tyler C. Shimko and Erik C. Andersen*. "Using natural variation to decipher the complex genetic cause of *C. elegans* drug sensitivities." 19th International *C. elegans* Meeting. June 2013
- 4) 4/2013 - Tyler C. Shimko*, Erik C. Andersen, and Leonid Kruglyak. "Identifying the genes that control paraquat resistance in the roundworm *C. elegans*." National Conference on Undergraduate Research.
- 5) 2/2013 - Tyler C. Shimko*, Erik C. Andersen, and Leonid Kruglyak. "Identifying the genes that control paraquat resistance in the roundworm *C. elegans*." Utah Conference on Undergraduate Research.
- 6) 4/2012 - Tyler C. Shimko*, Christian Frokjaer-Jensen, and Erik M. Jorgensen. "Universal Transgene Insertion in *C. elegans*." University of Utah Bioscience Symposium for Undergraduate Researchers.
- 7) 3/2012 - Tyler C. Shimko*, Christian Frokjaer-Jensen, and Erik M. Jorgensen. "Universal

GRE Scores

Every Fellowship applicant is REQUIRED to take at least the aptitude portion of the Graduate Record Examination and to have the results (Verbal, Quantitative and Writing or Analytical scores) sent to the Foundation. This can be done by adding the Hertz Foundation ID Number, 4366, to those of the schools requiring the GRE for admission as a graduate student on your GRE form. If you have already taken the GRE, you can enter your scores in the boxes provided. We continue to accept GRE scores after the application deadline. Initial interview decisions will be made by mid-November. If you enter scores, you must also send documented verification of these GRE scores to the Foundation. MCAT scores may be accepted in lieu of GRE scores if you are already enrolled in medical school.

	Score	%ile	Date of Exam (mm/yyyy)	GRE Subject 1	Subject 1 Score	%ile	Date of Exam (mm/yyyy)
Verbal	163	92	06 / 2014				/
Quantitative	161	80	06 / 2014	GRE Subject 2	Subject 2 Score		
AWA/Writing	5.0	93	06 / 2014				/
or MCAT Score			/				

Transcripts

You must submit 1 set of official transcripts of all your college academic records, undergraduate and (if applicable) graduate. Official transcripts are those submitted directly from the school to the Hertz Foundation. (You may collect these and send them, but be sure they are still in the sealed, signed envelopes, prepared by the school's Registrar.)

Please list here the schools you have attended, the dates you were enrolled, your GPA while attending each school, and any degrees earned. We will expect you to send an official transcript from each of these schools. (If you briefly attended an institution, e.g., for a summer course, or for a class that is unrelated to your major, e.g., a foreign language session, you do not need to furnish a transcript of this schooling.) Please indicate which of the schools in your list below will be providing transcripts by placing an "x" in the "Transcript ?" box on the right.

Transcripts 1	Enrollment dates. mm/yyyy	GPA	Degree earned	Transcript? X
University of Utah	08 / 2011	3.89	BS	X
Transcripts 2		/		
Transcripts 3		/		
Transcripts 4		/		
Transcripts 5		/		
Transcripts 6		/		
Transcripts 7		/		
Transcripts 8		/		

Other Opportunities

The Foundation is often approached by other institutions, including industrial organizations, universities and national laboratories, who are interested in considering our Applicants or Fellows for part-time or summer employment or in offering other types of opportunities. The Foundation also receives requests for information concerning Applicants from other funding organizations, both governmental and private charities, who are seeking well-qualified applicants to consider for fellowships. The Foundation will normally share information on a Fellowship application with potential employers and other funding organizations that the Foundation considers to be responsible, disinterested and acting in the public interest, provided that the Applicant has not asked that such information be withheld. If you would prefer that we NOT share the information in this application with these other organizations, please check the "Do NOT make other opportunities available" box below. (Your chances for selection as a Hertz Fellow are not influenced by this choice.)

Do NOT make other opportunities available to me:

Optional Survey

How did you hear about the Hertz Foundation Fellowship?

Referred by Professor

If you heard about the Hertz Foundation through a press release, news story, or other source, please use the box below to provide a brief description:

SUBMIT THIS

Were you a Goldwater Scholar or Nominee?

Choose one of the following:

Scholar

APPLICATION

ONLINE

Reference Reports

Please list here the names, addresses, telephone numbers and email addresses of the four persons who best know your academic and professional ability and whom you are requesting to provide Reference Reports on your behalf. Note that these Reports must be received by the Foundation by Monday, November 3, 2014 or your application is not assured of full consideration by the Foundation. IT IS YOUR RESPONSIBILITY TO ASSURE TIMELY RECEIPT OF ALL FOUR OF THESE REFERENCE REPORTS BY THE FOUNDATION. Once your referees are registered, they will receive an email from Embark explaining how to submit the Reference Report. This email will include the necessary access information and password.

Reference Report 1

title/position
school or business
address

phone
email

title	first name	last name
Dr.	Erik	Andersen
Assistant Professor		
Northwestern University		

Reference Report 2

title/position
school or business
address

phone
email

title	first name	last name
Dr.	Gillian	Stanfield
Assistant Professor		
University of Utah		

Reference Report 3

title/position
school or business
address

phone
email

title	first name	last name
Dr.	Leonid	Kruglyak
Professor		
University of California, Los Angeles		

Reference Report 4

title/position
school or business
address

phone
email

title	first name	last name
Dr.	Erik	Jorgensen
Professor		
University of Utah		

Contact Information

After examination of all applicants for the Fannie and John Hertz Foundation Graduate Fellowships, approximately 20% of the applicants will be offered interviews. These interviews are technical in nature, covering topics in physics, chemistry, mathematics, engineering, quantitative biology, etc.

Most interviewers are volunteers, fitting the meetings into their work or travel schedules. Interviews may take place at hotel conference rooms, schools, airports, restaurants, or other convenient locations. If we request that you travel more than five miles to meet with a Foundation interviewer, we will fully reimburse your reasonable-and-ordinary travel expenses to attend the interview. **However, please note the Hertz Foundation will only reimburse up to \$400 for airfare roundtrip from outside of the United States.**

Because the volunteers sandwich interviews into their other activities, there may be only two or three days notice of your interview. We must know how to let you know of a possible interview from **November 10, 2014 to March 15, 2015.**

In the boxes below, please enter telephone numbers where we can reach you, or an answering machine, or a friend, or a relative, during this time, so that you will be able to contact the Hertz Foundation within 24 hours of our attempt to reach you. Please identify each number, e.g., "my work #," "my parent's #," "my dad's # at work," "my advisor's #." We will try the numbers in the order you give them, but the more options you give us, the more likely it is that we will be able to reach you when we need to. Please also include email addresses, as we will be using email as our first line of communication in most cases.

()
SUBMIT THIS
()
APPLICATION
()
ONLINE
()

ADDITIONAL INFORMATION

Tyler

Shimko

36469464

I have benefitted tremendously from my intense pursuit of research opportunities. I recognize that, throughout my undergraduate career, these experiences have allowed me to grow as a person and a scientist. I have spent much of my free time advising others on the rewards and sense of accomplishment that scientific research can bring. As a researcher, I have experienced new cities and uncovered knowledge. Science has allowed me to see what was once invisible. I have fallen in love with scientific research and, as Carl Sagan once said, "When you're in love, you want to tell the world."

I was fortunate to gain access to academic research early at the University of Utah through the school's Office of Undergraduate Research. In order to provide my fellow students with similar opportunities, I joined the office as an Undergraduate Research Ambassador and Advisor. In these positions, I was able to talk with students and prospective students about the unique opportunities that research had afforded me and direct them to laboratories and research groups pertinent to their interests. I also helped to set up and oversee the university's annual Undergraduate Research Symposium, which afforded me the chance to get to know the projects of my fellow students. My experience with the Undergraduate Research Symposium taught me a valuable lesson about the importance of communication within the research community. I recognized that the most interesting and effective presenters were those who could accurately and succinctly describe their research to a non-expert. In order to improve my own communication skills, I sought out a position as a contributor to the Public Library of Science's (PLOS) Student Blog shortly thereafter. I realized that the Student Blog would be an excellent platform for me to experiment with different communication styles and to have an audience outside of my area of expertise, an opportunity not necessarily available through traditional academic publishing channels.

In my writing for the Student Blog, I have focused on developing, yet broadly applicable, topics within the biological sciences. For instance, my most recent post discussed the implications of increased experimental automation within biology, especially as it relates to molecular biology and the phenotyping of model organisms. I find that keeping the topics interesting, yet not overly technical, draws in members of the general public and piques their interest in my own research and that of my peers. The implication of this finding with respect to the broader public acceptance and trust in scientific endeavors does not escape me. I have learned a great deal about effective communication in my experience writing for the blog. I hope to one day apply these communication skills alongside my array of research-related skills as a professor and researcher at a major research university.

I maintain a commitment to openness and transparency in both my academic and professional development. I ensure that all software I write is publicly accessible and licensed in such a way that allows for free reuse and modification as well as the continued sharing of any descendant software. I publish in open access journals whenever financially possible and will make preprints of all manuscripts available in online repositories when I cannot. As I move forward in my academic career I will continue this commitment to open, accessible, and reproducible research. I believe that growth within the sciences can be fostered by lowering barriers to research not only in the way in which results are published, but also through outreach and introduction of the research process at an early stage in students' academic careers. My interest in outreach and the responsible representation of science has led to my participation in some unique activities outside of the realm of academic research.

During my sophomore year, I joined my roommate in creating a student group to encourage the university to divest its endowment from investments in the two hundred largest fossil fuel companies. Given the overpowering influence of these companies in the state of Utah, we understood that while we would face strong opposition, success in divestment would make a profound statement about the university's position on climatic issues. During the campaign, I utilized my skills in computer science and statistics to gather, analyze, and present information about student support for our proposed resolution. Ultimately, while our campaign did not succeed in causing divestment, our efforts were not in vain as the student body and university now have an open dialog about responsible energy consumption.



COPASutils: An R Package for Reading, Processing, and Visualizing Data from COPAS Large-Particle Flow Cytometers

Tyler C. Shimko, Erik C. Andersen*

Department of Molecular Biosciences, Northwestern University, Evanston, Illinois, United States of America

Abstract

The R package COPASutils provides a logical workflow for the reading, processing, and visualization of data obtained from the Union Biometrica Complex Object Parametric Analyzer and Sorter (COPAS) or the BioSorter large-particle flow cytometers. Data obtained from these powerful experimental platforms can be unwieldy, leading to difficulties in the ability to process and visualize the data using existing tools. Researchers studying small organisms, such as *Caenorhabditis elegans*, *Anopheles gambiae*, and *Danio rerio*, and using these devices will benefit from this streamlined and extensible R package. COPASutils offers a powerful suite of functions for the rapid processing and analysis of large high-throughput screening data sets.

Citation: Shimko TC, Andersen EC (Shimko,) COPASutils: An R Package for Reading, Processing, and Visualizing Data from COPAS Large-Particle Flow Cytometers. PLoS ONE 9(10): e111090. doi:10.1371/journal.pone.0111090

Editor: Denis Dupuy, Inserm U869, France

Received July 23, 2014; **Accepted** September 26, 2014; **Published** October 20, 2014

Copyright: © Shimko, Andersen. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. This R package is available through CRAN (<http://cran.r-project.org>). Newest updates are available through github (www.github.com/AndersenLab/COPASutils). The package and raw data can also be found at www.AndersenLab.org/Research/Software.

Funding: This work was supported by start-up funds provided by Northwestern University. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: erik.andersen@northwestern.edu

Introduction

High-throughput screening is an increasingly important task in many fields of biology [1]. The Complex Object Parametric Analyzer and Sorter (COPAS) platform and the BioSorter device from Union Biometrica allow for rapid and automated collection of flow data, including object length, optical density, and three fluorescence channels, from a large variety of organisms. Comprising five distinct machines, these devices can be used to analyze and sort organisms ranging in size from 10–1500 microns, including *Caenorhabditis elegans* [2], *Anopheles gambiae* [3], and *Danio rerio* [4]. Data are output in a flat, tab-delimited text file format, which is machine-readable. With the addition of an optional ReFLx module or LP sampler, samples are collected from a standard 96-well microtiter plate, allowing for well-by-well analysis. Although tab-delimited text files are read and understood by machines with existing software tools, they do not represent the best human-interpretable way to analyze and visualize 96-well data.

A large number of software environments can be used for analysis of data from these large particle measurement devices, including R, SAS, and MATLAB. The R environment offers several distinct advantages over other programming environments, including free and open-source software packages along with a vibrant community of scientists developing novel software. Two existing packages could be used to analyze COPAS data, but they are limited to the expensive MATLAB environment [5] or specifically written for cell culture RNAi experiments [6]. R is the standard for statistical computing and the fastest growing

environment for analysis of large data sets in computer science and biology.

We developed COPASutils, a novel software package for the R statistical programming environment to assist in the reading, processing, analysis, and visualization of data from 96-well plates, specifically data from the different COPAS platforms. COPASutils leverages the most recently developed R tools to rapidly read, summarize, and clean data to present elegant and adaptable visualizations of both summary and distribution statistics in a complete, compact, and intuitive pipeline structure. It represents a significant improvement over existing software packages by leveraging the R environment and research community to analyze data specifically generated using the powerful large-particle flow cytometers from Union Biometrica.

For users new to R

We recommend searching online resources for introductions to R. Tutorials for the interface, data handling, and scripting are the most useful introductions. The source code of the COPASutils package is available on the Comprehensive R Archive Network repository and can be installed by typing the following command in R: `install.packages("COPASutils")`. For a tutorial on the usage of COPASutils and the available functions, please see the vignette at <http://andersenlab.org/files/COPASutilsVignette.html>. This package is open-source. For updates and to submit comments, please go to <https://github.com/AndersenLab/COPASutils>.

Design and Implementation

COPASutils is designed to be simple to make it addressable to users familiar with the R environment. The functions included in the package have been constructed in such a way as to adapt to the data entered as input, so that users are not overwhelmed by options when attempting to use the functions. In the implementation of the COPASutils package, speed and extensibility are paramount. Data summarization functions are built upon the dplyr package [7] to minimize computational time and load. In addition, all plotting functions are built upon the popular ggplot2 visualization package [8] and, as such, resultant plots can be further modified to fit user specifications using standard ggplot2 grammar and functions. The complete source code and binaries for COPASutils are available from the Comprehensive R Archive Network (<http://cran.r-project.org/web/packages/COPASutils>). Because the entirety of the COPASutils package is written in the R language, this package is easily portable across platforms as well as completely free to utilize and modify. Overall, COPASutils presents a branched linear workflow for the analysis of COPAS data (Figure 1) that is simple for novice users of R as well as extensible for experienced users and users with specific experimental designs.

Data import and bubble detection

Data from a COPAS machine are output to tab-delimited text files. Each line of the text file represents the observation of one object that passed through the flow cell. However, the COPAS software cannot differentiate objects of interest from bubbles or particulate contamination. This issue is common for projects that require liquid handling with the bypass of the bubble trap hardware or for experiments where a chemical precipitate is present. Consequently, we have designed a function to read in the data that can eliminate many of these false-positive observations. The function required to read in the data can optionally set minimum and maximum cutoffs for the size and optical density of the objects, allowing the user to exclude data objects beyond the size parameters of the organism being measured. This function

works for data obtained from any flow cytometer with the ability to sample from 96-well plates, including the COPAS and BioSorter platforms.

Additionally, the function can optionally employ a support vector machine (SVM) [9] to probabilistically determine if an object recorded by the COPAS software is an undesirable bubble or a true object to be measured. This support vector machine must be trained on each COPAS or BioSorter device because internal parameters vary by device. To train the SVM, run two 96-well microtiter plates through your device. The first plate should contain only objects of interest with the bubble trap engaged, and the second plate should contain no objects of interest, less liquid per well, and have the bubble trap disengaged. These plates will generate object and bubble data, respectively. A variety of R packages (like [9]) can be used to generate an SVM for use with the readPlate function. With the combination of particulate contamination reduction and bubble detection and removal, the COPASutils package represents a powerful method of data filtering that greatly reduces the number of false objects analyzed.

The function used to read in plate data will also normalize the optical density and fluorescence values measured for every object. The device outputs integrated measures of each of these parameters. Therefore, larger objects always have increased optical density or fluorescence values. To better represent the optical density or fluorescence of individual objects, we divide each parameter by the length (time of flight) for every object recorded. This normalization allows the researcher to determine if objects of different size also have different optical density and fluorescence.

Summarization, well removal, and data filling

After the data are read in, it is often necessary to summarize the data by well in order understand patterns by strain, treatment, or replicate. Therefore, we have included a function to aid in the summarization of the data by well. The summarization function calculates not only means and variances of all distributions of measured parameters but also quantiles along with minimum and maximum values. Additionally, this function can also summarize

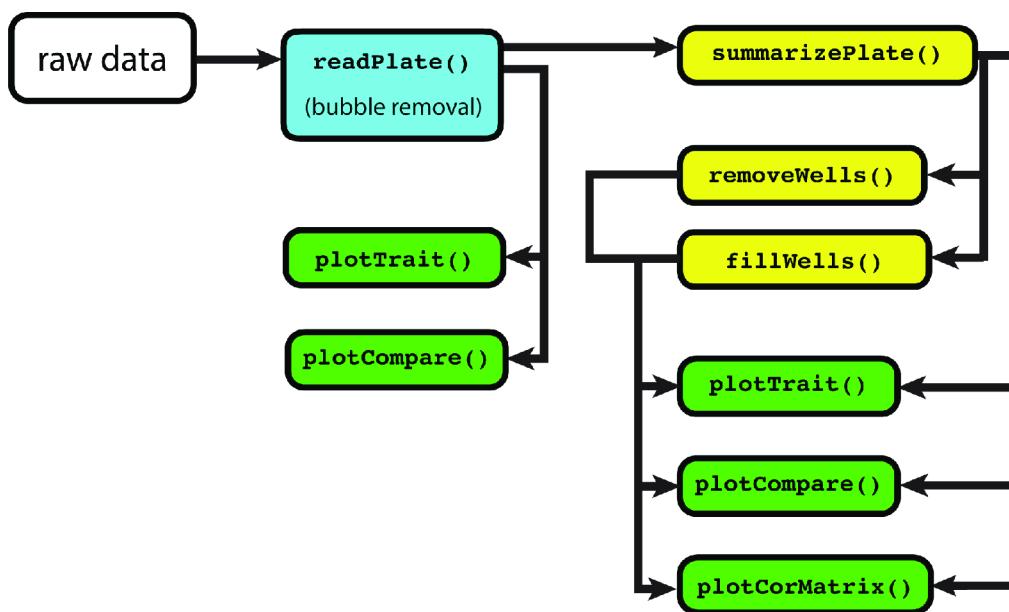


Figure 1. COPASutils workflow map. Suggested workflow for COPAS data using the COPASutils package. Reading steps are shown in blue, processing steps are shown in yellow, and plotting steps are shown in green.
doi:10.1371/journal.pone.0111090.g001

these statistical measures for log-transformed parameters. These values can describe the center, spread, and tails of the distributions for each feature, separated by each well of a microtiter plate.

The package also contains tools for further processing of the data after it has been read into R and summarized. COPAS or BioSorter machines with either a ReFLx module or LP sampler are capable of measuring object characteristics for every object from each well in a 96-well plate. However, some experimental designs do not require the use of all 96 wells. For this reason, the package includes a function for the removal of specific wells from the data set, either by setting all trait data from those wells to NA or by dropping those entries entirely. The function accepts a character vector of wells to be removed as input, allowing for well removal to be scripted and automated based on parameters of the summarized data.

Moreover, if wells are unused in a particular experimental design, the COPAS or BioSorter machines might fail to record any observations from the empty wells. This scenario may be problematic for the downstream steps of processing and visualizing the output data. To compensate for unused wells in the summarized data, a function has been constructed that will automatically detect missing wells and add the wells to the summarized data. The wells will be included in the summarized data with NAs for all of the population statistics, making further processing and visualization consistent.

Plotting and cross-plate comparisons

Because summarization might reduce the amount of data analyzed by the user, it is often best to visualize both the distribution of the original data and the values of the summary statistics to recognize patterns and identify potential errors. Included in the package are several plotting functions that make the visualization steps of the analysis simple and beautiful. The first of these plotting functions generates scatter plots and histograms for traits in a raw data frame (Figure 2a and 2b) or heatmap plots for traits in a summarized data frame (Figure 2c). This plotting function is particularly advantageous as it allows users to explore correlations between wells or between traits within the same well.

Correlations can also be explored on a much larger scale within the summarized data. We have included a separate function for the plotting of a correlation matrix between all traits in a summarized data frame either within a plate or between plates. Employing this function, users can determine which of the observed traits correlate with one another (Figure 3). For example, a user can recognize that the mean length of an organism is negatively correlated with population size, indicating that there could be competition for resources such as nutrients or oxygen. This type of relationship is experimentally relevant and immediately identifiable using a correlation matrix between traits.

Additionally, we have included a function to generate plots that allow the user to compare distributions (Figure 4a) or summary statistics across plates (Figure 4b). Distributions of raw data are represented as box and whisker figures in a boxplot. Similarly, summarized data are represented as bars in a bar plot. All of the individual plots are drawn in the position of their respective wells in a standard 96-well format. In this manner, data in the same well can be directly compared across plates. Even though plate-well series plots have previously been considered for this type of cross-plate comparison [10], the well-by-well plots presented in the COPASutils package represent a more straightforward representation of data which do not necessarily display any trend with increasing well number. Therefore, the plotting functions included in the COPASutils package are expected to have broader applicability than others currently available.

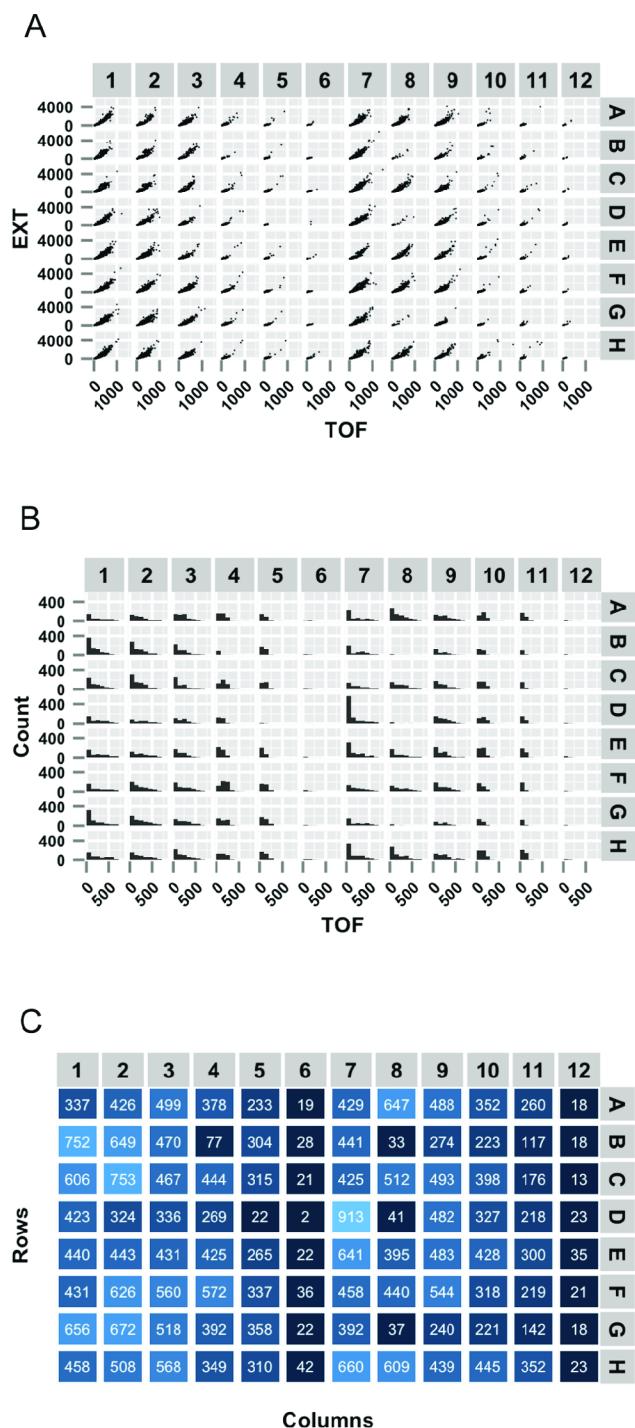


Figure 2. Example trait plots. Three possible plots made using the plotTrait function. (A) Well-wise scatter plots of the time-of-flight (TOF, a measure of length) values plotted against the extinction (EXT, a measure of optical density) values from raw data. (B) Histograms of the TOF values from raw data. (C) Heatmap of the population size from the summarized data (variable n) of each well. All data from this figure are included with the COPASutils package and deposited at <https://github.com/AndersenLab/COPASutils-data> and <http://andersenlab.org/Research/Software/>. doi:10.1371/journal.pone.0111090.g002

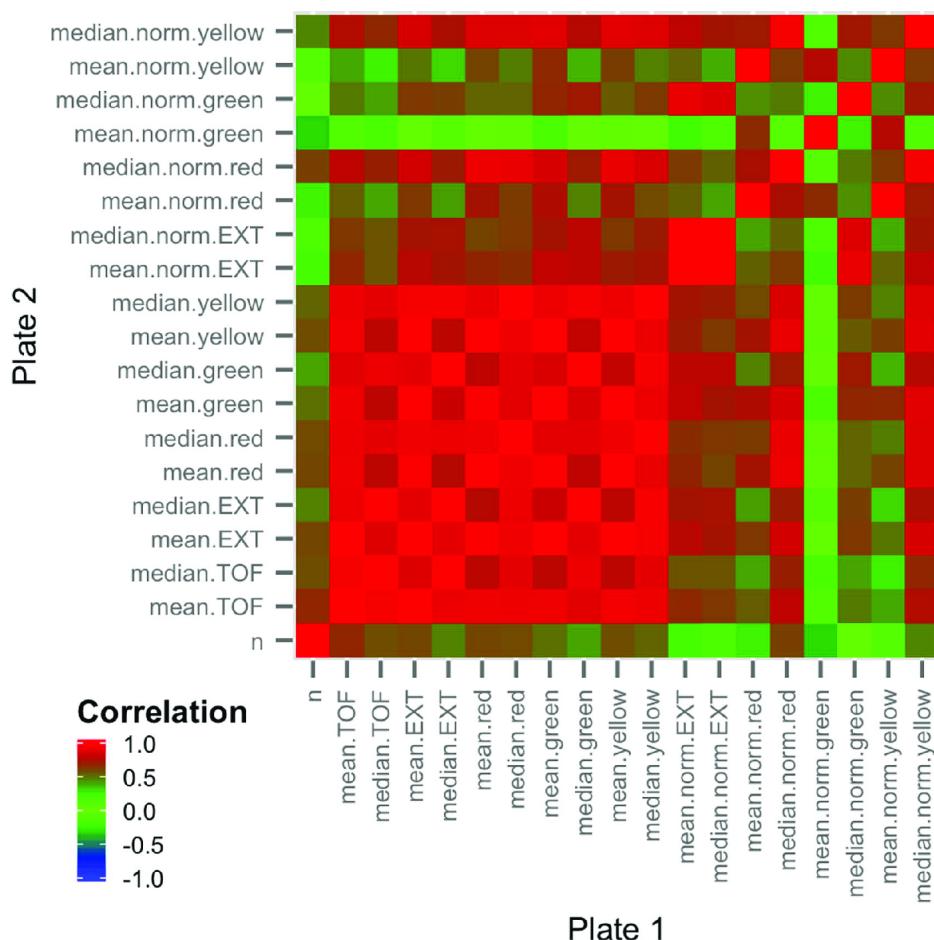


Figure 3. Example correlation matrix plot. A correlation matrix is plotted for all traits within a single plate using the `plotCorMatrix` function. Correlation matrix plots can be generated within a single plate (as shown) or between two plates. Positively correlated traits are shown in red, uncorrelated traits are shown in green, and negatively correlated traits are shown in blue. In this plot, all traits related to the mean and median values of optical properties of the objects (EXT, red, yellow, green) appear highly correlated, indicating that it may not be useful to include all such traits in further analysis.

doi:10.1371/journal.pone.0111090.g003

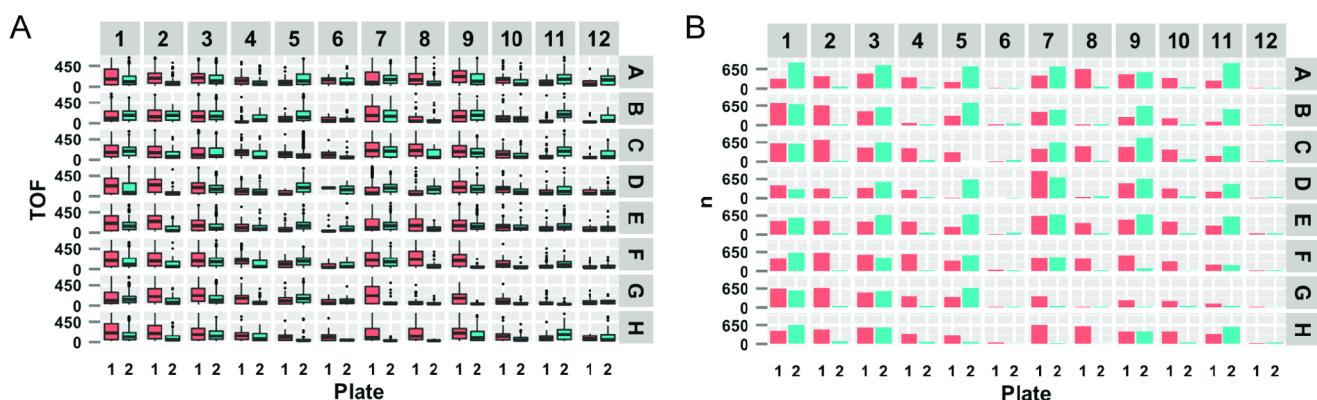


Figure 4. Example comparison plots. Two possible plots made using the `plotCompare` function. (A) Cross-plate comparisons, from raw data, of the distribution of the time of flight values within each well. (B) Comparison of population sizes, from summarized data, between two plates. All data from this figure are included with the `COPASutils` package and deposited at <https://github.com/AndersenLab/COPASutils-data> and <http://andersenlab.org/Research/Software/>.

doi:10.1371/journal.pone.0111090.g004

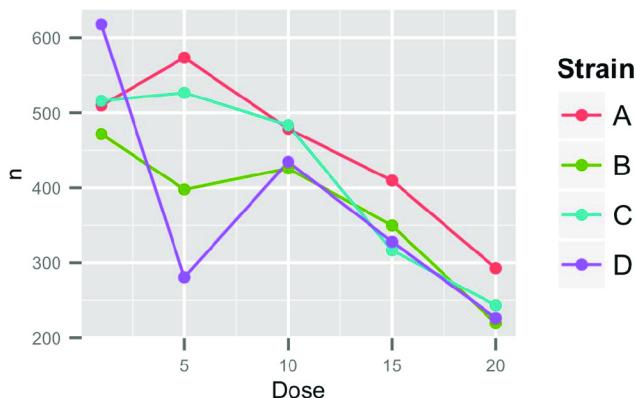


Figure 5. Example dose response plot. A dose response curve is plotted for four different strains in five different conditions using the `plotDR` function. Each point represents a single observation. Lines represent overall change in population size between doses. Dose response curves can be plotted for any trait present in the summarized data frame.

doi:10.1371/journal.pone.0111090.g005

Dose response experiments

COPAS machines are particularly well equipped to measure dose-dependent responses after exposure of animals or cells to increasing concentrations of drugs, RNAi, or other environmental perturbations. Because of the widespread utilization of the large-particle sorting devices for these experiments [11] functions have been included that allow the plotting of dose-response data (Figure 5). These functions require additional input from the user in the form of two 96-element vectors, one for the strains present in the plate, which is passed to the summarization function, and one for the dosages used, passed to the plotting function. The user then specifies the trait to be examined and a plot will be generated that summarizes the trait by strain and plots the mean of the selected trait with increasing dosage. A second function is also included that will return a list of plots for all traits present in the dataset, which allows for the plotting and analysis to be scripted.

Edge-effect detection

For long-running experiments, one can observe effects caused by the position of the well within the microtiter plate. These

References

- Giacomotto J, Ségalat L (2010) High-throughput screening and small animal models, where are we? *British Journal of Pharmacology*. doi: 10.1111/j.1476-5381.2010.00725.x.
- Boyd WA, McBride SJ, Rice JR, Snyder DW, Freedman JH, et al. (2010) A high-throughput method for assessing chemical toxicity using a *Caenorhabditis elegans* reproduction assay. *Toxicology and Applied Pharmacology*. doi: 10.1016/j.taap.2010.02.014.
- Bernardini F, Galizi R, Menichelli M, Papathanos P, Dritsou V, et al. (2014) Site-specific genetic engineering of the *Anopheles gambiae* Y chromosome. *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.1404996111.
- Veneman WJ, Stockhammer OW, de Boer L, Zaaij SAJ, Meijer AH (2013) A zebrafish high throughput screening system used for *Staphylococcus epidermidis* infection marker discovery. *BMC Genomics*. doi: 10.1186/1471-2164-14-255.
- Elizabeth M, Lamitina T (2010) A Suite of MATLAB-based Computational Tools for Automated Analysis of COPAS Biosort Data. *Biotechniques*. doi: 10.2144/000113427.
- Zhang XD, Zhang Z (2013) *displayHTS*: a R package for displaying data and results from high-throughput screening experiments. *Bioinformatics*. doi: 10.1093/bioinformatics/btt060.
- Wickham H, Romain F. (2014) *dplyr*: a grammar of data manipulation. R package version 0.2. Available: <http://cran.r-project.org/web/packages/dplyr/>.
- Wickham H. (2009) *ggplot2*: elegant graphics for data analysis. New York: Springer. 213 p.
- Karatzoglou A, Smola A, Hornik K, Zeileis A. (2004) *kernlab* - An S4 Package for Kernel Methods in R. *Journal of Statistical Software* 11(9), 1–20.
- Zhang XD, Yang XC, Chung N, Gates A, Stec E, et al. (2006) Robust statistical methods for hit selection in RNA interference high-throughput screening experiments. *Future Medicine*. doi: 10.2217/14622416.7.3.299.
- Boyd WA, Smith MV, Kissling GE, Freedman JH (2010) Medium- and high-throughput screening of neurotoxicants using *C. elegans*. *Neurotoxicology and Teratology*. doi: 10.1016/j.ntt.2008.12.004.

effects, caused by uneven oxygenation or evaporation, can obfuscate measurements of populations within wells of the plate. For instance, the outermost wells of a plate are most likely to experience the greatest evaporation, which might affect the outcome of a particular experiment. In order to test for these “edge-effects,” we have included a function that will split the plate into two populations, edge wells and center wells, and return the p -value(s) for a two-sided Wilcoxon Rank-Sum test for either a user-defined trait or for all traits. By utilizing this function, users can recognize plates that were affected by well position effects and remove those data from any further analyses. Users implementing the COPASutils package can ensure that their data are as clean and consistent as possible.

Conclusions

The COPASutils package provides simple reading, processing, and plotting tools for data obtained from the powerful COPAS platforms in the R statistical programming environment. The package presents an organized workflow for the management of COPAS data. By leveraging existing R tools, COPASutils offers rapid summarization and modular visualization design that is extensible to many unique projects. In addition, the package includes tools that offer compact and understandable visualizations for comparing data across wells, plates, or functions for early-stage detection of anomalous data. COPASutils is a simple, extensible, and reproducible open-source analysis pipeline for COPAS data.

Acknowledgments

The authors thank Union Biometrica for providing sample data from the BioSorter machine with LP Sampler and the members of the Andersen Laboratory for feedback and suggestions while developing COPASutils.

Author Contributions

Conceived and designed the experiments: TCS ECA. Performed the experiments: TCS ECA. Analyzed the data: TCS ECA. Contributed reagents/materials/analysis tools: TCS ECA. Contributed to the writing of the manuscript: TCS ECA. Designed the software in this report: TCS ECA.