A personal narrative of 6 pre-university research projects over 7 years (2009-2015) yielding 19 manuscripts

Abstract

Acquisition of research skills, including scientific enquiry, is an important requirement in scientific education, after the acquisition of a body of fundamental knowledge. Working on research projects is a direct means to gain research skills, as well as gaining a firsthand experience of the research environment. Here, I shall narrate my experience and learnings as a research mentor involving 22 pre-university students in 6 research and development projects over a period of 7 years, yielding a total of 19 peer-reviewed manuscripts. I have 3 intentions to this narration - (a) to demonstrate that pre-university students can carry out useful research, (b) summarize my learning experience in this journey, and (c) providing some pointers and encouragement to my fellow mentors and intended mentors. I learnt that (1) pre-university students can produce publishable work but (2) the scope of the projects must be well-defined with specific and measurable end-points, (3) the involvement of the mentor is substantial both in project formation and project management, (4) quality work can be achieved when students understand the background and context of their work, (5) enduring working relationship between students and mentors requires time and efforts to build, and (6) the students can pleasantly surprise the mentor with their quality of writing and reasoning. Upon reflection, I feel that I gained as much as the mentees/protégés, if not more, and this encouraged me.

Keywords: pre-university, high school, research projects, publications, critical success factors

Introduction

Scientific enquiry is important in scientific education,\(^1\) in addition to acquiring a body of knowledge. Upon acquiring a body of knowledge, the student is expected to be able to apply the knowledge in new settings. A common tool for teaching scientific enquiry is to engage students in project work or problem-based learning;\(^2\) which had been carried out at all levels of scientific education, from high school\(^3,4\) to undergraduate studies.\(^5\)\(^6\) Such research skills are exceptionally crucial in bioinformatics as bioinformatics is primarily a hands-on, minds-on endeavour.\(^7\) This emphasis on bioinformatics research and problem-solving skills are supported by significant number of bioinformaticians.\(^8\)\(^9\)\(^10\)\(^11\)

Several studies have enumerated various benefits of incorporating research experience into scientific education.\(^5,8,9,10,13,15\) Krasny\(^5\) studied high-ability students and found that students gained a better understanding of the research process while graduate students acting as mentors gained their first mentoring experience. Moreover, students with research experience are more likely to enter\(^8,13\) and maintain a career in science.\(^6\) This may be due to increased motivation, independence, and active learning,\(^2\) which may be the result of increased confidence and role-modeling of the research mentor.\(^11\) Research experience is crucial in providing a platform to develop data handling and analytical skills, and gaining a deeper understanding of the theoretical principles.\(^15\)

In this article, I account my experience as a research mentor over 7 years (2009-2015) using a multi-case\(^7-20\) personal narration\(^21-25\) approach involving 22 pre-university students in 6 research and development projects with several intentions- (a) to demonstrate that pre-university students can carry out useful research, (b) summarize my learning experience in this journey, and (c) providing some pointers and encouragement to my fellow mentors and intended mentors. Personal narration, sometimes also known as storytelling,\(^21\) is autobiographic in nature with rich emotional and contextual cues\(^22\) that are required to describe the human experience.\(^23\) Thus, personal narration puts the human into the story,\(^26,27\) Personal narration also overlaps with participatory action research where reflective self-inquiry is carried out with the intention to improve subsequent actions.\(^28-30\) In a series of cases over a period, it is highly unlikely that learning from preceding cases does not impact on succeeding cases even though this is not intentional. Despite so, there are elements of participatory action research spanning across my narration. Multi-case study is an extension of single case study\(^21,22\) where more than one cases are presented. Thus, multi-case study offers the ability to compare between different cases\(^21,23\) to unravel underlying patterns while providing the rich and in-depth contextual analysis in single case studies.\(^31,32\)

Personal narration

In this section, I will narrate the 6 research and development projects and within each project, the narrative will be in chronological order to illustrate the learning process and progression within each project. A timeline and publication achievements of the 6 projects and 22 students in question is given in Figure 1 below.
A personal narrative of 6 pre-university research projects over 7 years (2009-2015) yielding 19 manuscripts

<table>
<thead>
<tr>
<th>Project</th>
<th>Student</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Development of CyNote</td>
<td>YY Ng</td>
<td>2009</td>
</tr>
<tr>
<td></td>
<td>CY Chia</td>
<td>2010</td>
</tr>
<tr>
<td></td>
<td>WT Leong</td>
<td>2010</td>
</tr>
<tr>
<td></td>
<td>CWX Lim</td>
<td>2010</td>
</tr>
<tr>
<td></td>
<td>EJG Ang</td>
<td>2011</td>
</tr>
<tr>
<td></td>
<td>QH Chu</td>
<td>2011</td>
</tr>
<tr>
<td></td>
<td>YJ Lin</td>
<td>2011</td>
</tr>
<tr>
<td></td>
<td>OYW Chan</td>
<td>2011</td>
</tr>
<tr>
<td></td>
<td>SSJ Heng</td>
<td>2011</td>
</tr>
<tr>
<td></td>
<td>BMH Keng</td>
<td>2012</td>
</tr>
<tr>
<td>Finding Reference Genes</td>
<td>CY Chia</td>
<td>2012</td>
</tr>
<tr>
<td></td>
<td>WT Leong</td>
<td>2012</td>
</tr>
<tr>
<td></td>
<td>CWX Lim</td>
<td>2012</td>
</tr>
<tr>
<td></td>
<td>EJG Ang</td>
<td>2012</td>
</tr>
<tr>
<td></td>
<td>QH Chu</td>
<td>2012</td>
</tr>
<tr>
<td></td>
<td>YJ Lin</td>
<td>2012</td>
</tr>
<tr>
<td></td>
<td>OYW Chan</td>
<td>2012</td>
</tr>
<tr>
<td></td>
<td>SSJ Heng</td>
<td>2012</td>
</tr>
<tr>
<td></td>
<td>BMH Keng</td>
<td>2013</td>
</tr>
<tr>
<td>Evolutionary Context of Codon Usage Bias</td>
<td>CY Chia</td>
<td>2013</td>
</tr>
<tr>
<td></td>
<td>WT Leong</td>
<td>2013</td>
</tr>
<tr>
<td></td>
<td>CWX Lim</td>
<td>2013</td>
</tr>
<tr>
<td></td>
<td>EJG Ang</td>
<td>2013</td>
</tr>
<tr>
<td></td>
<td>QH Chu</td>
<td>2013</td>
</tr>
<tr>
<td></td>
<td>YJ Lin</td>
<td>2013</td>
</tr>
<tr>
<td></td>
<td>OYW Chan</td>
<td>2013</td>
</tr>
<tr>
<td></td>
<td>SSJ Heng</td>
<td>2013</td>
</tr>
<tr>
<td></td>
<td>BMH Keng</td>
<td>2014</td>
</tr>
<tr>
<td>Experimental Evolution: Food Additives</td>
<td>CH Lee</td>
<td>2014</td>
</tr>
<tr>
<td></td>
<td>KC Lee</td>
<td>2014</td>
</tr>
<tr>
<td></td>
<td>JSH Oon</td>
<td>2014</td>
</tr>
<tr>
<td></td>
<td>WC Ng</td>
<td>2014</td>
</tr>
<tr>
<td></td>
<td>ZE Chay</td>
<td>2014</td>
</tr>
<tr>
<td></td>
<td>ZQ Aw</td>
<td>2014</td>
</tr>
<tr>
<td></td>
<td>BZL Loo</td>
<td>2014</td>
</tr>
<tr>
<td></td>
<td>SXZ Low</td>
<td>2014</td>
</tr>
<tr>
<td></td>
<td>DJW Goh</td>
<td>2014</td>
</tr>
<tr>
<td>Experimental Evolution: Halophilization</td>
<td>JA How</td>
<td>2015</td>
</tr>
<tr>
<td></td>
<td>JZR Lim</td>
<td>2015</td>
</tr>
<tr>
<td></td>
<td>CFG Castillo</td>
<td>2015</td>
</tr>
<tr>
<td></td>
<td>ZE Chay</td>
<td>2015</td>
</tr>
</tbody>
</table>

Figure 1 Timeline of projects.

Development of electronic notebook (CyNote)

My doctoral thesis is on bioinformatics. Instead of using regular paper-based laboratory notebooks, I used an electronic laboratory notebook (ELN), jointly developed by Oak Ridge National Laboratory, Lawrence Berkeley National Laboratory, and Pacific Northwest National Laboratory, under U.S. Department of Energy’s DOE2000 Collaboratory Program; to manage my scripts and analysis output. I envisioned an ELN for biologists and bioinformaticists where common bioinformatics tools can be made available via the ELN and the analysis results be stored. I have been using Python programming language for all my work since my honours and postgraduate years; hence, a default programming language to base my subsequent work on.

The opportunity came when I was a lecturer in the School of Chemical and Life Sciences (CLS) at the Singapore Polytechnic, and was approached by a colleague in School of Media and Information Technology to consider accepting an internship student, YYng. A semester long internship is a mandatory requirement for the diploma that YYng was pursuing. The director of CLS approved of this internship.

YYng worked in the library made it a point to learn Python programming. At that point in time, my intention was to develop a web-enabled ELN rather than a standalone system as many of the bioinformatics functions provided by BioPython, a Python library for computational biology, requires web access. Our original intention was to develop on CherryPy, a Python web application framework, but quickly found that CherryPy is too “low leveled”, which is useful for complete customization. Web2Py, on the other hand, made many assumptions and provide a set of application structure (even though in-depth customization is still possible) to enable rapid application prototyping, which is suitable for CyNote development. I met up with YYng every morning in the library to discuss the issues he faced the preceding day and development for the day. YYng was encouraged to ask for help in Web2Py Google Group where I will supplement his questions in the group. Massimo, the creator of Web2Py, and other active members of the group are patient to our questions and rendered significant help. Once YYng realized that an ELN is essentially a
stricter version of a blog application (the common ground), together with the application framework of Web2Py, YYng made great strides in development – where notebook entries are synonymous with blog entries and supplementary notations on each entry are synonymous to comment entries within a blog entry. With each progress and solved issue, YYng’s confidence and enthusiasm increased. At the end of his internship, we have a working CyNote where I can add bioinformatics tools on. Subsequently, we managed to present CyNote as a 21 CFR Part 11 compliant ELN in the inaugural PyCon Asia Pacific as a peer reviewed conference paper, with YYng as co-author. YYng passed his repeated subject, awarded “A” for his internship (which significantly improved his grade point average as internship is equivalent to one semester of coursework), and a peer-reviewed article in his name.

Finding reference genes

This project started as a project under the Science Mentorship Programme (SMP, www.moe.gov.sg/education/programmes/gifted-education-programme/special-programmes/science mentorship-programme), managed by the Gifted Education Branch at the Ministry of Education, Singapore. Its aim is to provide a platform for “developing pupils’ interest and talent in scientific research” where “they provide opportunities for pupils to deepen their knowledge in a particular area, to learn about the latest research developments, and to be challenged by the intellectually stimulating process of scientific research.” Only year 3s and 4s (15 to 16 years old) students in gifted education and recommended by their school, can participate in SMP as project students. A group of 2 or 3 students is assigned to a project instead of mammalian organs. As BMH Keng recounted their project formation years later:

“I decided to narrow into finding reference genes in microarray datasets from NCBI Gene Expression Omnibus. Every gene in each of the 8 datasets was analyzed for their coefficient of variation (CV), which is defined as the quotient of the standard deviation to its arithmetic mean. Hence, low CV corresponds to expression stability; thus, suitable as reference genes. After analyzing each dataset separately, we found that Microtubule affinity-regulating kinase 3 (MARK3) to have low CV (CV<0.08) across all 8 datasets, making MAPK3 the most suitable reference gene for mouse liver. More importantly, we found that commonly used reference genes, such as Glyceraldehyde-3-phosphate dehydrogenase (GAPDH), showed large expression variation (CV=0.47) in one of the datasets, making it unsuitable as a reference gene for mouse liver. A lot of time was devoted into writing their term paper as I saw it as consolidation of their understanding. I also witnessed examples of theoretical misunderstanding or uncertainties as they were tasked to put their thoughts and arguments on paper; thereby, providing me with a chance to correct them. As a result, their SMP term paper was submitted to IUBMB Life and accepted for publication,” which encouraged me continue my participation in this programme and further strengthened my belief that even high school students are able to contribute to the scientific community provided that the project is scoped appropriately. I was also awarded Outstanding Mentor by the Ministry of Education.

In the following year (2010), I proposed to find reference genes in mouse endocrine glands. This project was taken up by QH Chu, EJG Ang, and YJ Lin; 15-year old students from Raffles Institution. In this project, 15 microarray datasets were used and analyzed using the same method as Chia et al. Similarly, we found that GAPDH showed large expression variations (CV as high as 0.98), making GAPDH probably one of the worst choices for reference genes, but what was more interesting was that MARK3 that was found to be a suitable reference gene in mouse liver was found to be within the list of suitable mouse endocrine reference genes. In addition, this study reflected the findings of Chia et al.,48 that ribosomal proteins are relatively stable in expression. However, the biggest drawback of this study is that the microarray datasets in question are from a variety of microarray chipset, resulting in difficulties in a proper comparative study as different microarray chipset can have different sets of probes. Hence, this project is not suitable for publication in its own right even though they are required to write up a term paper. Nonetheless, this project forms a suitable case example to show that common housekeeping genes, such as GAPDH, may not be suitable reference genes in every case. This is supported by studies showing that GAPDH has stable expression profile in reticulocytes but varies in expression across different developmental stages of rat fetal brain.47 Thus, this project forms one of the case studies in a peer-reviewed book chapter on reference genes; thereby, generating a publication out of their work. Looking back, this project would be substantially more rigorous should datasets from the same microarray chipset were used and we realized the problem too late into the project. This also taught me that the mentor must consider the project more thoroughly, including running through the entire project mentally, as pre-university students will not have the experience to anticipate such issues.

I left Singapore Polytechnic at the end of 2010 and through a series of event; I got my final SMP group in 2011. They were 15-year old students from Raffles Institution whom had been unsuccessful in securing a SMP project. Their school acted upon this to run their SMP as school-based option in which the school had to either find an external mentor or assign one of their teachers as project mentor. Due to my involvement in QH Chu’s SMP, I was asked if I was interested to take on a project which I agreed. This enabled a flexible working schedule and we will meet up every weekend or alternate weekends at a café to discuss project progress and to clear any doubts. In this project, I took the lesson from QH Chu’s project and ensured that I had all the datasets from the same microarray chipset. In addition, I decided to narrow into finding reference genes in Escherichia coli instead of mammalian organs. As BMH Keng recounted their project formation years later:

“I remember we failed to get any of the SMP projects we applied for during the first two rounds of project selection, which was pretty disappointing. Fortunately, our teacher-mentor managed to get hold of a few more options for us, of which the E. coli project was one. I thought it sounded good because it seemed to have the most relevance to biology, which I was interested in. The workload also didn’t seem too demanding, which was nice because we were fairly heavily involved in core curricular activities and other activities at the time. So we came to a consensus that the E. coli project was the most appealing.

It was early 2011 that I had a meeting with Dr. Too Heng Phon (Department of Biochemistry, National University of Singapore) and...
he made a casual comment on Chia et al., saying that CV may not be stringent enough. Hence, in this new project with Heng et al., we used CV as the first screening. From the 4 datasets obtained from NCBI GEO, 3 datasets came from E. coli MG1655 and 1 came from E. coli W3110. We treated the datasets separately—10% CV was used as threshold to potential reference genes from each of the E. coli MG1655 datasets and 39 genes were common between all 3 potential gene lists, which were then subjected to NormFinder. We found that Glucan biosynthesis protein G (mDgo) is a suitable reference gene for both MG1655 and W3110 strains of E. coli, and found that GAPDH showed high expression variation and was not in the potential list of reference genes. Their SMP term paper was published almost in verbatim and years later, both BMH Keng and OYW Chan felt that it was still an important step in their scientific pursuits—“Looking back, I’d say that publishing made this research experience stand out from my other science-related pursuits” and “I definitely think the publication of that first paper was not to be overlooked.”

After the successful completion of their SMP, Heng et al. were interested to move onto the next project even though I had taken up a position in South Dakota State University, USA. At the end of the E. coli project, I was surprised that if it is unlikely for an organism-specific reference gene (a reference gene suitable for various organs within the same organism), can it be possible for organ-specific reference gene (a reference gene for a specific organ across various organisms)? Two microarray datasets, from the livers of Spermophilus lateralis (ground-roaming golden-mantle ground squirrel) and Spermophilus tridecemlineatus (tree habiting 13-lined ground squirrel) each, under different states of feeding and hibernation had been studied using the same microarray chip, were found in NCBI GEO. The experimental conditions were the same in these 2 datasets, making them a valuable resource for examining the possibility of organ-specific reference genes; or more specifically in this case, genus-organ or Spermophilus liver specific reference genes. Due to geographical and time zone differences (USA vs Singapore), progress was not as soon as desired— as OYW Chan recounted:

“We did emailing and instant messaging to keep in touch with Maurice, but some clarifications are just best done over pen and paper in person. We asked many dumb questions multiple times, which Maurice was patient with, but was clear evidence of the lowered efficiency. We never thought of abandoning it though. We were sure that we’d complete it eventually - it was just a question of how much progress we could discipline ourselves into accomplishing thatweek.”

However, I will strongly attribute this to the determination of the group. Eventually, our results suggested that Spermophilus liver specific reference genes is not likely as suitable reference genes in S. lateralis liver are not suitable for S. tridecemlineatus liver and vice versa. Since S. lateralis and S. tridecemlineatus are from the same genus and the livers are harvested at the same feeding conditions; thereby, enabling rigorous comparison which is the intention of Williams et al., we extrapolated to genus-organ-specific reference genes are not likely. By further extension, it suggests that lineagespecific reference genes are not likely, and there is a constant need to identify reference genes for each organ and in each species. These findings are published. In a certain sense, this is a negative result as we failed to identify Spermophilus liver specific reference genes but it presented a good case that negative results can be published if the project scope is clear and the scientific rigors are maintained, as BMH Keng recounted—“At first I thought it quite weird that we were publishing negative results, so the process was quite eye-opening for me.”

One of the issues with available tools for finding reference genes from expression datasets; such as, NormFinder, geNorm, and BestKeeper; is the inability to cope with large datasets containing thousands of genes, as in the case of microarray datasets. For example, geNorm can only process 102 genes. This required transcriptome sized datasets to be filtered into candidate reference genes before analyses by these tools. Hence, I was considering the development of a method that is comparable to these tools but with the ability to handle transcriptome sized datasets. Therefore, I proposed the idea of a method development project to BMH Keng and OYW Chan in 2013 (SSJ Heng opted to leave the group after the spermophilus project), which excited them, as BMH Keng recounted—“The fact that we had succeeded in publishing two papers gave us confidence in starting another one. I think we had gained a much better understanding of the field at the time, plus the premise of the third project sounded quite interesting - not just using available tools to find reference genes like the first two projects, but actually developing our own algorithm”. We selected a subset of microarray data for analysis using NormFinder, geNorm, and BestKeeper. The outputs were correlated—with the correlation between NormFinder’s output and geNorm’s output being the highest (r=0.987) while the coefficient of determination (r’) between NormFinder/BestKeeper and geNorm/BestKeeper being 0.489 and 0.459 respectively. These results are consistent with study showing that the identified reference genes by NormFinder, geNorm, and BestKeeper vary even on the same dataset. However, these coefficients of determination set a minimum correlation between our algorithm(s)—we need to achieve a coefficient of determination of at least 0.489 when correlated to NormFinder’s output. This was an exploratory project and OYW Chan and BMH Keng were encouraged to use their creativity to come up with new methods by building on existing methods; hence, it was of great fun. Eventually, we found 2 algorithms that achieved coefficient of determination of 0.892 when correlated to NormFinder’s output. In addition, these 2 algorithms perform linearly to the number of data points. For example, if 10 samples of 1000 genes (10,000 data points) took 2 seconds to complete processing, 100 samples of 5000 genes (500,000 data points) is estimated to take 100 seconds to process. Two publications, one publishing the results and algorithm and the other publishing the tool, resulted from this work. For the sake of ownership and fun, we had decided to name the tool as OLIVER (OLigonucleotide Variable Expression Ranker), after OYW Chan.

Evolutionary context of codon usage bias

By the end of OLIVER, OYW Chan and BMH Keng were in the middle of their first year in a Junior College (equivalent to year 11 and 12) and were due to take their General Certificate of Education (Advanced Level) at the end of the following year. Hence, I was not expecting a fourth project but they were enthusiastic about it, as BMH Keng recounted the scene then—“Well I guess we were just talking about it at Coffeebean, and somebody said “ok why not” and the other person said “yeah let’s do it, I’m sure we can manage”. I don’t think we really pondered too much into it.” I was surprised and did not expect it. As a result, I was almost scrambling to think of a new project. At that point in time, I had just completed a NIH-funded project in South Dakota State University, examining evolution of antisense transcripts. Hence, I have a deep interest in applying statistical genomics into the study of evolution, and since my previous employment as a Senior Scientist in Life Technologies, I had nurtured an interest in codon usage bias (CUB). After a few days of mental brewing, I decided to propose a project to examine the null hypothesis that CUB is not evolutionarily conserved, using the same statistical
techniques used in my antisense expression study. Using 18 codon usage tables across both eukaryotic and prokaryotic species from Codon Usage Database (www.kazusa.or.jp/codon), we found that CUBs is correlated within species of the same class and the CUBs between closely related organisms are higher correlated compared to more distant organisms, suggesting that CUBs corroborates evolutionary time. This is the same conclusion from my antisense expression study and that of Jordan et al. In addition, we showed that visualizing CUBs on a dendrogram fits our currently known phylogeny and we published this work in Asia Pacific Journal of Life Sciences.

This was the last project I did with OYW Chan and BMH Keng as they had to prepare for their milestone examination the following year. In total, they had worked with me for 4 years and generated 5 peer-reviewed journal articles and 1 peer-reviewed book chapter. In preparation for this narrative, I did ask them if they would go through all these all over again and would they encourage their juniors to do the same – their responses are positive.

**Bryan:** Overall, I enjoyed the process of working on this series of papers, and if given the opportunity, I would do it again. Doing multiple interrelated projects in the same field really helped to solidify my understanding and made me better able to assume greater ownership of the work we did. I think the process was a lot more meaningful, in terms of personal growth as a researcher, as compared to completing just a single project. …… I feel that completing multiple projects is useful in honing the skills necessary to reach the eventual publication of a manuscript.

**Oliver:** This research experience taught me a great deal about the way science is done and the rationale behind it. As a student, I learnt much about a specific set of bioinformatics fields and as a growing adult I learnt much about myself. Student research provides opportunities that are uncommonly replicated elsewhere – independent work, depth of inquiry, exposure to a field expert. ……. Yes. Please embark on a project. Don’t stop at one paper because a good working relationship takes time to build and so does a base of usable field knowledge.

**Experimental evolution: food additives**

As a lecturer in the Singapore Polytechnic, I was required to take 1 or 2 final year projects (FYP) with 3 diploma final year students (about 19 years old) per project. As Diploma in Biotechnology (a 3-year full time course) emphasized on practical skills (by the time each student enters final year, they would have accumulated about 300 hours of laboratory training, excluding 8 weeks of full time internship), there was an unspoken rule that most FYPs have substantial experimental component. Having completed my honours project and doctoral thesis in the Department of Zoology (The University of Melbourne, Australia), the importance of evolution was engrained into me, as Oliver: This research experience taught me a great deal about the way science is done and the rationale behind it. As a student, I learnt much about a specific set of bioinformatics fields and as a growing adult I learnt much about myself. Student research provides opportunities that are uncommonly replicated elsewhere – independent work, depth of inquiry, exposure to a field expert. ……. Yes. Please embark on a project. Don’t stop at one paper because a good working relationship takes time to build and so does a base of usable field knowledge.

I spent a significant amount of my time with my project group, often having lunch and dinner together. I am blessed with students who have the time availability (financially sound and not having to work extensively to make ends meet) and enthusiasm to be willing to stay back after class to discuss their project, analyze the data, and write drafts after drafts of their report. I had the group start drafting their project report, starting from literature review, barely 3 months into their year-long project. I had chased KC Lee for his draft for nearly a month when CH Lee and JSH Oon were onto their second or third draft. Eventually, KC Lee submitted his draft to me and to my surprise; it was well-written and only required minor edits. I had seen KC Lee’s former laboratory reports before and the draft he gave me after more than a month was miles better. This project was tedious but in the end, we found that *E. coli* can adapt to various additive media as generation time decreased over passages. However, our results suggest
that the cells adapted to their specific media as swapping the media resulted in increased generation time. This is even for the case where cells in high concentration were swapped into low concentration of the same additive. DNA fingerprinting suggests mutations were involved in adaptation; yet, the fingerprints converged over passages suggesting that there is a presence of global stress response which had been suggested by previous studies.44,45 The amount of work sustained over extended period resulted in 2 peer-reviewed publications46,47 from this project.

This project was continued for another academic year with a new group of students–BZL Loo, QZ Aw, and SXZ Low – extending the sub-culture from passage 70 to 150. The transition was relatively easy as JSH Oon had kindly guided the new group before he was due for his mandatory national service and the protocol had been developed. In fact, BZL Loo, QZ Aw, and SXZ Low continued the laboratory notebook that CH Lee, JSH Oon, and KC Lee used; which enabled them to have a good sense of the data logging requirements. My personal guidance had been consistent – we had a formal meeting every Friday, when generation time were taken, where we discussed the progress of the project and resolve any issues faced. I recalled that I enjoyed the process of guiding the group to compare between their results and that of the previous group.46 This work resulted in a succeeding manuscript48 where two interesting observations emerged. Firstly, the rates of adaptation (in terms of decline in generation time) were reduced compared to pre-70th passage,7 suggesting that the adaptations were approaching plateau. I taught the group to estimate 95% confidence interval of their linear regression (decline in generation time across passages) using Jack-knife re-sampling. Secondly, the cells adapted to different concentrations of the same chemical additives as though they were different additives rather than different concentrations of the same additives, suggesting possibility of adaptation into an ecological niche (also known as ecological specialization). This was supported by DNA fingerprinting suggesting that the cells diverged genetically.

Near to the end of BZL Loo, QZ Aw, and SXZ Low’s FYP, I was wondering whether there will be any worth continuing this adaptation study as it was becoming obvious that the rate of interesting observations is likely to decrease. I had also thought that should I carry on this project for the next academic year, I might look at loss of adaptation instead. However, that was the point of time that I left the Singapore Polytechnic.

In addition, two side projects had emerged out of this adaptation project. As I was searching for measures to measure genetic distance based on DNA fingerprinting, I came across a list of other potential distance measures and I had implemented most, if not all, of them into a Python module.46 Subsequently, another student, ZE Chay (from the same batch as CH Lee), approached me indicating that he wanted to learn Python programming and work on something while waiting for his time for enlistment into national service (mandatory conscription for all Singaporean males). I got him to learn Python programming online and given my interests in statistics his first project (a practice project) was to implement statistical testing routines in Python, which resulted in a code paper.46 Upon gaining more confidence with Python programming, I got ZE Chay to examine 20 other distance measures for their applicability in DNA fingerprinting using the DNA fingerprinting data from CH Lee, JSH Oon, and KC Lee’s FYP where Nei and Li’s distance measure7 was used. Nei and Li’s distance67 is identical to Dice’s distance.88 We found that certain distance measures; such as Simpson’s distance,89 have negligible correlation (r=0.008) with Dice’s distance;88 hence, not suitable as substitute for Dice’s distance.88 Russel and Rao’s distance measure,90 on the other hand, is highly correlated (r=0.998) with Dice’s distance;88 thus a suitable substitute for Dice’s distance.88 A manuscript was published on these findings.48

Another side project originated at the early phases of the main adaptation project. One of the key “amazement” of Lenski’s long-term experimental evolution is the number of generations that E. coli had undergone11–40 and measuring the number of generations is crucial for this very purpose. We used optical density as an indirect measure of cell density. However, Sezonov et al.,91 studied E. coli K-12 in Luria-Bertani broth (LB) and found that the average cell mass decreases after optical density (OD 600) more than 0.3. After which, a decreasing growth rate was observed, which the authors91 attributed the depletion of fermentable sugars. However, LB is able to sustain extended periods of cell growth but at a slower growth rate due to the high availability of amino acids, which the cells could utilize as a carbon source. This resulted in the need to correct for decreased cell size when optical density is used as proxy for measuring cell numbers, such as in this project. However, this raised two questions. Firstly, is depletion of sugars the cause of reduced cell growth in other common media; such as, nutrient broth (NB); which was used in this project? Secondly, if there is decreased cell mass at high optical density, is it possible to use DNA content as a proxy to estimate cell numbers?

We performed media exhaustion experiment by repeated harvesting and re-inoculation of fresh cultures into LB, NB, and brain heart infusion (BHI), for 11 weeks and found that the sustainable cell counts (by spread plate counting) declined steadily to plateau by 8th week after peaking at 4th week. The pH of the media were neutral after 11 weeks (7.1<pH <7.7). However, addition of glucose into LB, NB, or BHI rescued the media; thereby, confirming that lack of sugar is the major reason for growth decline.92 We also found low correlations between optical density and viable plate counts (r<0.2), suggesting a change in cell size.93 Using the genomic sequence of E. coli ATCC 8739, the genomic DNA content per million cells can be estimated at 515.35 nanograms (ng). By extracting genomic DNA from different aliquots of cells with known cell density and quantifying the extracted DNA, we extracted an average of 527ng of genomic DNA per million cells, with a standard error of 10.605ng, which is not significantly different from the estimated genomic DNA content. With the amount of extracted genomic DNA, we are able to estimate the number of cells and this is highly correlated (r=0.93) to the results from viable plate count. Hence, we demonstrated that lack of sugars results in growth depletion in LB, NB, and BHI; and cell size changes at different cell densities; and DNA content can be a proxy to estimate cell numbers; thus, publishing a manuscript in the process.93 The attraction of this side project is the time contributed is minor compared to the amount of time contributed to the main project. In fact, this was considered low priority project that we nearly dumped. Looking back, I am glad that we persisted on.

**Experimental evolution: halophilization**

This project stemmed from the first food additive project by CH Lee et al. At that point in time, I was quite certain that E. coli is able to adapt to various environmental conditions should sufficient time be given. Hence, I was considering a FYP project to stress the cells. A few options were considered; including, increased temperature, reduced oxygen, and increased preservatives (such as, benzoic acid). However, all these choices were not feasible due to resource constraints. For example, the laboratory for running FYPs is also

---

**Citation:** Ling MHT. A personal narrative of 6 pre-university research projects over 7 years (2009-2015) yielding 19 manuscripts. MOJ Proteomics Bioinform. 2017.6(3):270–281. DOI: 10.15406/mojpb.2017.06.00193
of Desmond and Jian Ann, Samuel [SXZ Low] / Bryan [BZL Loo] didn't manage to get their balloted option of Mr Goh, I saw it as an opportunity to bring everyone together by suggestion and vouching for Dr Ling to them, and I recall Desmond aiding me in trying to pull them in with us”.

JZR Lim, JA How, and DJW Goh were coached by JSH Oon and WCng during the first 2 weeks of experiments. This project involved regular sub-culturing and generation time estimation; hence, we planned it to coincide with ZQ Aw, SXZ Low, and BZL Loo’s project so that both teams can work together. It was also helpful that SXZ Low, BZL Loo, JA How, and DJW Goh were in the same class; hence, identical lesson schedule. This also meant that JZR Lim and ZQ Aw could be backup for experiments if required. The same arrangement happened with CH Lee, JSH Oon, and KC Lee as well, where CH Lee and JSH Oon were classmates and KC Lee acted as backup for experiments. As a result, project discussions were usually done together across both groups.

Over the course of project, we managed to culture *E. coli* to 8% NaCl, which was above bacteriostatic level suggested by Doudoroff.92 Despite so, minimum inhibitory concentration assay demonstrated that 8% NaCl adapted *E. coli* could grow in concentrations as high as 11% NaCl. Comparatively, 11% NaCl is half the saturation level of NaCl in water and about 3 times the salinity of sea water (average NaCl concentration in sea water is 3.5%). Hence, we had reduced the water activity requirement of *E. coli*. We were concerned about *Staphylococcus aureus* contamination as *S. aureus* is commonly found on human skin and is known to be highly salt tolerant.93,94 To ensure that *S. aureus* contamination was kept at bay, we cultured the cells on MacConkey agar before cryopreserved as glycerol stocks to ensure integrity of the glycerol stocks. We performed Gram staining regularly and used DNA fingerprinting to assay for contaminations. I recalled that we had to revert a few passages from glycerol stocks due to contaminations. All four replicates showed an increasing rate of generation time as the concentration of NaCl increases, suggesting osmotic stress to the cells. Hence, there was a need to balance between culture survival and the rate of NaCl increase. As the cultures adapted to increasing NaCl concentrations, we wondered whether the cultures varied their salt optimality (increasingly able to grow in higher NaCl concentrations while reducing the ability to grow in lower NaCl concentrations) or broadening the range of salt tolerance (increasingly able to grow in higher NaCl concentrations while maintaining the ability to grow in lower NaCl concentrations). Interestingly, one of the cultures appeared to be broadening the range of salt tolerance while the other three cultures appeared to be varying their salt optimality. We published these findings.92,93

**Evolution by digital organism**

I left the Singapore Polytechnic in December 2010 and one of the issues was the custodian of the glycerol stocks that I had left behind. Although I managed to hand-over the custodial duties to the school (approved by the director of the school), I was left wondering how much enthusiasm I will have if I am asked to restart this experimental evolution study. Unless I am certain of extended employment or if there was sufficient resource for third-party custodian, I would find it difficult to restart knowing the possibility of not being able to bring these glycerol stocks with me. This was the point where I wondered and started to examine alternative platforms for my evolutionary studies. One of the possibilities was to study evolution in situ, which was to observe natural evolution, much like Charles Darwin and Alfred Wallace. In this case, glycerol stocking is not logistically possible
as there is no possible way to freeze an island for the purpose of preservation. Another possibility was to run computer simulations of evolution using artificial life (ALife) organisms, also known as digital organisms (DOs).55 DOs or ALife are based on cellular automata, described by Christopher Langton,68 and had been used to study many aspects of evolution.97–103 Studies using DOs had been supported by evolutionary biologist; such as, Thomas Ray, whom wrote one of the earliest artificial life simulators,98 and even well-known experimental evolutionary biology; such as, Richard Lenski; had used DOs in his studies.101–104 In fact, Richard Lenski co-authored Avida, a well-known artificial life simulators, with Charles Ofría.101 The premise of studying evolution using DOs is intriguing to me compared to observing natural evolution. Essentially, working with DOs amount to creating one or more populations to test evolutionary hypotheses or events, which can be difficult to perform experimentally.105 At the same time, it fits my interests in bio-inspired algorithms; such as, genetic algorithms;106 and resolves the issue with glycerol stocks as “glycerol stocks” in DO context are computer files, which can be duplicated and stored at various locations at ease. I decided to write a DO library for Python, Digital Organism Simulation Environment (DOSE)107, 108 during my leisure time in South Dakota, USA.

As the Vice-President of Python User Group (Singapore), I was involved in the organization of PyCon Asia-Pacific (APAC) Conference since its inception in 2010. During the conference days of the Third PyCon APAC, I was introduced to Clarence Castillo, whom was a student volunteer from Republic Polytechnic, at the conference. Clarence was pursuing a course in Diploma in Information Technology then. While chatting with the student volunteers, I found that Clarence showed strong enthusiasm and interest in problem solving; hence, I was toying the possibility of taking Clarence as my protégé. I was surprised that Clarence has an underlying interest in biology and saw the uniqueness of using computer simulations to study evolution.

My original implementation of DOSE107,108 was rather rudimentary on hindsight and needed code refactoring, as well as feature enhancements. Clarence happily took up this challenge and painstakingly took each line of my code to understand its operations. The goal was to simplify the use of DOSE. The main enhancement was to add a 3-dimensional “world structure” onto the DOs. This enabled movements of DOs within the world and could be used to generate multiple habitats or homing ranges, which was not possible in a “world-less” population. After a few months of work, DOSE was finally ready for use.

We tested our product using a case study to examine the effects of migration on population genetic diversity. Three migration schemes were tested – in case one, the DOs remained in their own habitat (no migration); in case two, DOs can migrate to adjacent habitats (adjacent migration); and in case three, DOs can migrate across habitats (long migration). No migration can represent isolated tribes or villages. Adjacent migration can represent nomadic movements while long migration resembles transcontinental migrations observed today. The social aspect to this work would be that in the present day where transcontinental mating and resettlement are commonplace, are we turning into a global village genetically? Our simulation results where transcontinental mating and resettlement are commonplace, are today. The social aspect to this work would be that in the present day while long migration resembles transcontinental migrations observed

My reflections

As I was narrating my stories, I had a chance to re-examine the circumstances of each project and its outcomes. More importantly, gave me a chance to ask myself several important questions – (1) why I decided to be a project mentor, especially mentoring several projects that are not mandatory within my paid job scope (in fact, only the 3 FYPs are mandatory within my job scope); (2) what did I learn and would I have done it differently; and (3) given a second chance, would I have gone down this path again?

Looking back, I am hard-pressed to pinpoint a moment when I decided to be a mentor. However, I can list three factors that I consider to be instrumental. Firstly, I was the eldest in my generation (which means that I do not have any elder siblings and I did not have any elder cousins, both maternal and paternal) until I was 21 years old. What happened when I was 21 is a family matter, which is out of the scope of this article. Being the eldest in my generation predisposed myself to be a natural de facto mentor (or child leader) to several of my younger cousins; thereby, entrusting me with mentorship opportunities – more like, I was never given the opportunity to reject. Several studies105,110
suggested that previous mentoring experiences are indicative of the person’s willingness to mentor. Moreover, willingness to mentor is correlated to the number of mentoring experiences.117 Secondly, I believe that my past experiences as a protégé played a significant role. The most extensive mentorship experience was during my postgraduate studies where I had the great privilege to be working with good thesis advisers, Drs. Kevin Nicholas and Christophe Lefèvre, during my doctoral days. Both are old enough to be my father and they allowed me extensive latitude to explore my work while gently guiding my research. I could still remember the day when Kevin, who is also my honours project supervisor, told me that I was to start my PhD as a student but to end as a colleague; and if I had never argued with him during my postgraduate, it would be his failure as an adviser. Both Allen et al.,116 and Ragins & Scandura117 pointed out that positive experiences as protégé can be indicative of willingness to mentor. Perhaps it is also due to my birth order, I may be more willing to mentor from an intuitive point of view than logical reasoning of cost-benefits.118 If I were to be pressed for a reason, I might only be able to say that “it just feels like it is a right and worthy thing to do, and I am willing to do it”.

As I was writing these narratives, and re-reading them, a scene from Sex and the City (Season 5 Episode 7 – The Big Journey) popped into my mind – where Carrie Bradshaw met up with Mr. Big after her book reading and Mr. Big told Carrie, “I had no idea I hurt you so much. I got to say it was tough seeing it in print …… Not all in one sitting, one reading, one right after another, bam… bam… bam…” The same happened to my head – these events were scattered across more than a thousand pages of my diary and I had to simplify and put them into a structure as I put the events in writing; thereby, enabling me to see themes across various stories; which had been therapeutic for me.119–122

I realized that all the students in these narratives did not have an idea what they wanted to do in terms of the nature of the project in the beginning but rather, they wanted or had to do something. I attributed this to the general lack of experience on the part of the students; hence, it is the mentor’s duty to moot a project of sufficient content to interest yet not overwhelming to the students. It is also the mentor’s duty to contextualize the project in the bigger realm of science. My students showed interests to want to do something; such as, taking up my project offer out of their course requirements or coming to me with the intention to do something; or willing to do something extra and on top of what is needed for them; such as, reading up more or drafting a manuscript. In hindsight, this is about all they can offer and it can be a challenge for them to stick their heads out for a mandatory research project as there is a perceived expectation of the project on the part of the students even though the mentor may not see it that way.123 Despite so, there is a perceived burden on the students to meet the supposed expectations of the mentor as failure to meet such expectations can be distressing.124–126 It then becomes the mentor’s task to have to bridge this expectation gap and be extremely patient for students will ask the same stuffs multiple times. In a way, I did felt that the amount of effort to teach a semester and the amount of effort to mentor a project group is the same despite the number of students being significantly different – it is concentrating the attention needed by a lecture room of students onto just a team of 3 project students and it will not be just imparting of skills and knowledge. At the very initial stage, the role of the mentor is the advocate of the field where he should build interest and context of the work. Only then, can there be a stable path for knowledge and skills transfer. If mentoring is a graduate course, then knowledge/ skills transfer is a hurdle requirement. The mentor should be a friend, a counsellor, and likely, even a surrogate parent. I was told by a former colleague that there was a student once that got into trouble and was detained by the police, and the student called up his project mentor instead of his parents. In a way, mentoring relationship is trust building between the mentor and the protégé.120–122 Once the trust is built, it can be fulfilling and Ragins & Scandura123 had suggested that the willingness of mentors can be associated to previous positive experiences as a mentor.

Given a second chance, would I have gone down this path again? My answer will be a resounding yes. At an autistic level, being a mentor had given me substantial satisfaction123 that I had contributed significantly to someone’s learning and the contributing to my own explorations in science. Being asked by students is a great way for me to improve and consolidate my understanding in the field.116 By explaining to my students, I gained insights into how I learn and the inherent blind spots to my knowledge. Nothing beats the motivation to flip open a textbook or to search for another publication just to support your arguments or to clear your protégés doubts. At the external level, I had seen my protégés growth in their intellect and reasoning skills. More concretely, there were multiple instances where my protégés rode on their published papers and gained entries into their desired course of studies and awarded competitive scholarships. I believe that their success is one of the best returns for my efforts.129–130 Each of my mentors played an irreplaceable role in my achievements, now and in future. Therefore, why will I walk this path? Perhaps the simplest answer is – to repay the gratitude I have for my own mentors.

Acknowledgements
I wish to thank all my students for making all these possible, and my mentors for providing me with positive experiences.

Conflict of interest
The author declares no conflict of interest.

References
6. Coker JS, Davies E. Involvement of plant biologists in undergraduate and high school student research. Journal of Natural Resources and Life Sciences Education. 2002;31:44–47.

Citation: Ling MHT. A personal narrative of 6 pre-university research projects over 7 years (2009-2015) yielding 19 manuscripts. MOJ Proteomics Bioinform. 2017;6(3):270–281. DOI: 10.15406/mojibin.2017.06.00193


A personal narrative of 6 pre-university research projects over 7 years (2009-2015) yielding 19 manuscripts.