

Mastery Rubric for Bioinformatics (MR-Bi)

Reference: Tractenberg RE, Lindvall JM, Attwood TK, Via A (2019) The Mastery Rubric for Bioinformatics: A tool to support design and evaluation of career-spanning education and training. PLOS ONE 14(11): e0225256. <https://doi.org/10.1371/journal.pone.0225256>

In it, 14 criteria and 5 levels of each criteria are defined; with the 5 levels being

- **Novice (e.g., early undergraduate/new to bioinformatics), Bloom's 1:** remember, understand. Novices can engage with well-defined problems, with known solutions.
- **Beginner (e.g., late undergraduate, early Master's), Bloom's 2–3:** understand and apply. Beginners may use, but not choose, tools; they can engage with well-defined problems and apply what they are told to apply; the answers may not be known, and the Beginner would stop once this was apparent.
- **Apprentice (e.g., Master's, early doctoral student), Bloom's 3–4, early 5:** choose and apply techniques to problems that have been defined (either jointly or by others). The Apprentice can analyze and interpret appropriate data, identify basic limitations, conceptualize a need for next steps, and contextualize results with extant literature.
- **Early Journeyman (J1) (e.g., late doctoral student or just after graduation), Bloom's 5, early 6:** begin to evaluate (review) and synthesize novel life-science knowledge, and to develop abilities to integrate bioinformatics into research practice, with some mentorship. The J1 Journeyman can contribute to problem formulation, shows earliest establishment of independent expertise in the specific life-science area, and can confidently integrate current bioinformatics technology into that area.
- **Late/advanced Journeyman (J2) (e.g., doctorate holder), Bloom's 5, late 6:** expertly evaluate (review) and synthesize novel life-science knowledge, and integrate bioinformatics into research practice. The J2 Journeyman is independent and expert in a specific life-science area, and can select, apply and develop new methods. The J2 Journeyman formulates problems, considers the relevance of “what works” within this area to other life-science domains, so as to be an adaptable and creative scientific innovator without having to reinvent every wheel.

Hence, this rubric can be used as a professional development tool to see where each of us stand and where to develop next.

Criteria 1: General description of bioinformatics practitioner

- **Novice:** Reads, generally understands, but does not question, life science research (results). Beginning to recognize that “facts” are actually just the best-currently-supported theory. Limited engagement with uncertainty associated with “facts”; developing understanding of experimental design paradigms in biology, & own specific area of study.
- **Beginner:** Consolidates reading & understanding, beginning to learn how to analyze given biology problems (with software). Growing recognition that “facts” are typically the best-currently-supported theory. Engaging consistently with uncertainty associated with “facts”; deepening understanding of experimental design paradigms in biology, & own specific area of study.
- **Apprentice:** Reads & understands; reliably identifies methods (software & programming) for given problems. Chooses & executes correct analysis, not necessarily able to identify several methods that could be equally viable, depending on given research objectives. Qualified as a fluent, but not as an independent, scientist

who uses bioinformatics as a tool, but does not yet synthesize technology with biology to generate new research problems.

- **J1 Journeyman:** Qualified as an independent scientist who uses bioinformatics methodologies as part of routine practice. Poses novel scientific questions, & identifies data & technology to align appropriate statistical/analytical methods to desired scientific objectives. Experienced reviewer of relevant technical features of available bioinformatics methods. Newly independent expert in integrating bioinformatics technology/techniques into novel research problems in their area of expertise.
- **J2 Journeyman:** Independent scientist who expertly integrates bioinformatics & more traditional methodologies, as needed, to achieve desired objectives & contribute to the body of knowledge. Expert reviewer of relevant technical features of available bioinformatics options.

Evaluation	J2 Journeyman
Justification	<ul style="list-style-type: none"> • Independent scientists that routinely start new honours and masters level projects • Uses novel computational methods to evaluate and reconcile contradictions from experimental settings resulting in peer-reviewed manuscripts; such as, Castillo, CFG, Ling, MHT. 2014. Resistant Traits in Digital Organisms Do Not Revert Preselection Status despite Extended Deselection: Implications to Microbial Antibiotics Resistance. BioMed Research International 2014, Article ID 648389. https://pubmed.ncbi.nlm.nih.gov/24977157/ • Member of the National Supercomputing Centre Singapore (NSCC) Ad-Hoc COVID-19 Project Selection Committee • Reviewer of many journals of which, I am not an editorial board member (as of December 31, 2002) <ol style="list-style-type: none"> 1. Advances and Applications in Bioinformatics and Chemistry (ISSN 1178-6949) 2. Artificial Intelligence Review (Print ISSN 0269-2821, Online ISSN 1573-7462) 3. Algorithms (ISSN 1999-4893) 4. Cognitive Computation (COGN, Print ISSN 1866-9956, Online ISSN 1866-9964) 5. Computation (ISSN 2079-3197) 6. Engineering Biology (ISSN 2398-6182) 7. Foods (ISSN 2304-8158) 8. Genes (ISSN 2073-4425) 9. IET Software (ISSN 1751-8814) 10. International Journal of Molecular Sciences (IJMS, ISSN 1422-0067) 11. International Journal of Environmental Research and Public Health (IJERPH, ISSN 1660-4601) 12. Journal of Clinical Medicine (JCM, ISSN 2077-0383) 13. Machine Learning and Knowledge Extraction (MAKE, ISSN 2504-4990) 14. Materials (ISSN 1996-1944) 15. Mathematical and Computational Applications (MCA, ISSN 2297-8747)

	16. Molecules (ISSN 1420-3049) 17. Biotechnology & Biotechnological Equipment (Print ISSN 1310-2818, Online ISSN 1314-3530)
Development Plan	<ul style="list-style-type: none"> Continue my practice.

Criteria 2: Considerations for evidence of performance at this level

- **Novice:** Bloom's 1, early 2: remember, understand. Problems the Novice can engage with are well-defined, with solutions already known. Work does not generally reflect self-assessment.
- **Beginner:** Bloom's 2–3: understand & apply, but only what they are told to apply. Problems the Beginner can engage with are well-defined. Work reflects some self-assessment, when directed to do so.
- **Apprentice:** Bloom's 3–4, early 5: choose & apply techniques to problems that have been defined (either jointly or by others). Can analyze & interpret appropriate data, identify basic limitations & conceptualize a need for next steps/contextualization of results with extant literature. Seeks guidance to improve self-assessment of own work.
- **J1 Journeyman:** Bloom's 5, early 6: evaluate (review) & synthesize novel life science knowledge while developing abilities to integrate bioinformatics into research practice. Shows independent expertise in a specific life-science area, & confidently integrates current bioinformatics technology into that area. Beginning to critically evaluate experimental paradigms & their results, without knowing/ requiring that there be "one right answer". Consistently self-assesses own work.
- **J2 Journeyman:** Bloom's 6: prepared for independent scientific work. Expert in design & critical evaluation of experimental paradigms & their results. Self-assesses in own work, & encourages others to develop this skill.

Evaluation	J2 Journeyman
Justification	<ul style="list-style-type: none"> Supervised students from high-school level to graduate level on bioinformatics projects. To date (December 1, 2020), 3 publications from high-school projects, 22 publications from pre-undergraduate projects, and 19 publications from undergraduate projects; all as senior author and peer-reviewed. Engages in self-reflection of practice <ul style="list-style-type: none"> Ling, MHT. 2017. A Personal Narrative of 6 Pre-University Research Projects Over 7 Years (2009-2015) Yielding 19 Manuscripts. <i>MOJ Proteomics & Bioinformatics</i> 6(3): 00193. Ling, MHT. 2019. Science/Education Portraits IV: Experiences from a Decade as Informal Career Counsellor can be Summarized as Personopreneurship. <i>Acta Scientific Medical Sciences</i> 3(3): 151-156.
Development Plan	<ul style="list-style-type: none"> Continue my practice.

Criteria 3: Ethical practice

- **Novice:** Exhibits respect for community standards/rules for public behavior & personal interaction.

- **Beginner:** Learning how to recognize, & manifest respect for, intellectual property, professional accountability, & scientific contributions.
- **Apprentice:** Learning to recognize “misconduct” in the scientific sense. Learning to avoid, & respond to, misconduct; & the importance of neither condoning nor promoting it.
- **J1 Journeyman:** Learning the principles of ethical professional & scientific conduct. Seeks guidance to strengthen applications of these principles in own practice. Learning how to respond to unethical practice.
- **J2 Journeyman:** Practices bioinformatics in an ethical way, & does not promote or tolerate any type of professional or scientific misconduct. Seeks guidance in how/when to take appropriate action when aware of unethical practices by others. Practices, & encourages all others to practice, bioinformatics in an ethical way. Does not promote or tolerate any type of professional or scientific misconduct. Takes appropriate action when aware of unethical practices by others.

Evaluation	J1 Journeyman
Justification	<ul style="list-style-type: none"> • Enforces ethical practices in bioinformatics by making as much of the data sets related to publications publicly available - https://github.com/mauriceling/mauriceling.github.io/wiki/Data-Set-Downloads • However, not fully confident to be able to take appropriate actions when aware of unethical practices by others, should the situation occur (has not happened yet).
Development Plan	<ul style="list-style-type: none"> • Needs professional training to handle unethical practices by others, especially of colleagues.

Criteria 4: Prerequisite knowledge – biology (includes statistical inference & experimental design considerations)

- **Novice:** Basic knowledge of biology; little to-no awareness of the uncertainty inherent in experimental designs common in the life sciences. Thinking about the life sciences is based on uncritical acceptance of information as “factual” or “true”.
- **Beginner:** Advanced knowledge of biology, & basic knowledge of key bioinformatics methods. Very simple statistics/programs are run to answer pre-defined scientific questions. Learning to understand the uncertainty inherent in the scientific method, questions assumptions in the data & their relevance for given scientific problems (which arise from others).
- **Apprentice:** Thinking about life sciences integrates both experimental & bioinformatics/technological sources for data & knowledge. Understands the uncertainty inherent in the scientific method, questions assumptions in the data & their relevance for given scientific problems (which typically arise from others, or with others). Experimental design & statistical inference are recognized & exploited with guidance, to answer given scientific problems. Can recognize inconsistencies in biological data/experiments that are identified by others, but cannot troubleshoot experimental methods independently.
- **J1 Journeyman:** Recognizes the importance of, & is able to critically evaluate, the relevant literature, & understands historical background of the relevant biological system(s). Sufficient knowledge of a biological system(s) to be able to draw functional conclusions from analytical results. Collaborates with experts to inform the

next stages in the experimental design process (validating results, follow-up analyses, etc.).

- **J2 Journeyman:** Makes predictions to inform next stages of experimental design process. Evaluates relevant experimental methods that can be applied in any problem. Can generalize to other biological systems; independently solves biological problems that are innovative & move the field forward.

Evaluation	J1 Journeyman
Justification	<ul style="list-style-type: none"> • Taught experimental design and statistical methods at undergraduate and graduate level. • Able to direct laboratory work across state during the preparation of Ling, MHT, Ban, Y, Wen, H, Wang, SM, Ge, SX. 2013. Conserved Expression of Natural Antisense Transcripts in Mammals. BMC Genomics 14(1): 243. • Versed with fundamental molecular biology skills but not updated with the latest techniques.
Development Plan	<ul style="list-style-type: none"> • Continue my practice. • Keep up to date with latest biological experimental techniques.

Criteria 5: Prerequisite knowledge – computational methods (includes statistical inference & experimental design considerations)

- **Novice:** Basic knowledge of computational methods; little-to-no awareness of the relevance of computational methods for life sciences. No awareness of experimental designs or how these can be used or implemented in computational applications. Thinking about tools, computers, software, & programming is strictly unidimensional: i.e., extrapolation &/or abstraction of knowledge about computational methods to other systems, programs, or problems, are not possible. Can run software or execute code they are given (as appropriate) with precise instructions; cannot write a script or debug/troubleshoot.
- **Beginner:** Computers, software, tools, & programming are understood to be options for scientific work. Learning how to write & test code, run software, or use tools, as appropriate. Is developing awareness of the variety of bioinformatics tools, designs, & resources, but is not able to choose or apply the most appropriate of these for any given question; when choices are made, tools are used uncritically. Developing awareness that computational tools require input parameters, but uses the default settings. Learning to read, understand, troubleshoot, & make minor modifications to existing code/scripts. Does not synthesize results or outputs.
- **Apprentice:** Learning to test software & programming approaches to different types of problem. Experimental design & statistical inference using computing & algorithms are recognized & applied, with guidance, to answer given scientific problems. Learning “best practices” for programming, if programming is part of the task. Can write basic code in a given language or run appropriate software, using judgement, but not inventing or innovating. Cannot troubleshoot complex computational methods – will ask for guidance. Exploring alternatives to default input parameters across computational tools. Can apply knowledge of tools to interpret their results & output. Seeks guidance in synthesis of results or outputs.
- **J1 Journeyman:** Recognizes the importance of, is able to critically evaluate, & understands historical background of the relevant data, databases, algorithms, tools,

data analysis/statistical methods & computational resources. Can utilise these & justify trade-offs across methodologies (e.g., which statistical test to apply & what computational methods to use). Collaboratively synthesizes & critically questions analysis results & output from tools. Recognises the iterative nature of experiments (e.g., bench, data analysis, back to bench). Can write code/use tools to accomplish these, but collaborates with domain experts for identifying & articulating biological problems that are innovative & move the field forward.

- **J2 Journeyman:** Develops robust, well-documented, optimized, reproducible code &/or uses tools to address biological problems; moves away from standard procedures & innovates to accommodate new data types, tools, & techniques as needed. Can generalize to new coding languages or software/tools/resources.

Evaluation	J2 Journeyman
Justification	<ul style="list-style-type: none"> • Developed 3 important open-sourced bioinformatics tools /suite, which had been used routinely by my students and others <ul style="list-style-type: none"> ○ AdvanceSyn Toolkit (https://github.com/mauriceling/advancesyntoolkit; Forked 21 times and starred 6 times as of December 1, 2020) ○ Bactome (https://github.com/mauriceling/bactome; Forked 28 times and starred 3 times as of December 1, 2020) ○ DOSE (https://github.com/mauriceling/dose; Forked 27 times and starred 40 times as of December 1, 2020) • Each function / method in these codes are systematically documented and publicly available for scrutiny.
Development Plan	<ul style="list-style-type: none"> • Continue in the efforts to enhance computational tool offerings.

Criteria 6: Integrate interdisciplinarity

- **Novice:** Does not recognize life sciences as requiring integration of both experimental & computational/modeling approaches. Perceives disciplines as separate; integration only occurs when/as directed. Information, ideas & tools that are inter-disciplinary are used without question.
- **Beginner:** Beginning to think about life sciences as requiring integration of both experimental & computational/modelling approaches. Recognizes that interdisciplinarity is needed, but does not know how (or when) to do it, & requires direction. Learning the integrating process; learning strengths & weaknesses of biological & computational methods, but not sufficient to question assumptions from these & other disciplines.
- **Apprentice:** Understands that life sciences integrate both experimental & computational/modeling approaches; seeks guidance about how & when to integrate. Developing an understanding of the strengths & weaknesses of biological & computational methods, beginning to question fundamental assumptions from these & other disciplines for any given scientific problem (which typically arises from others, or in conjunction with others).
- **J1 Journeyman:** Collaboratively integrates across relevant disciplines to address, & solve, innovative biological problems. Tests multiple avenues to triangulate solutions, with minimal guidance. Recognizes the roles of interdisciplinary teams in the research

process, & the importance of integrating interdisciplinarity early on. Works effectively on interdisciplinary teams with minimal guidance.

- **J2 Journeyman:** Formulates innovative biological problems that require interdisciplinary solutions. Integrates methods & results to derive & contextualize solutions to biological problems. Consistently tests multiple avenues to triangulate solutions, while exploiting relevant findings from other disciplines. Actively builds interdisciplinary teams, as needed.

Evaluation	J2 Journeyman
Justification	<ul style="list-style-type: none"> • Actively involved with interdisciplinary teams (engineering / science) for work in synthetic biology. • Participate in teams across geographical zones, Singapore / USA / China, during my time in Life Technologies under synthetic biology group. • Worked with overseas collaborators for publications <ul style="list-style-type: none"> ○ Malaysia - https://pubmed.ncbi.nlm.nih.gov/30809243/ ○ India - https://pubmed.ncbi.nlm.nih.gov/22588998/ ○ Taiwan - https://pubmed.ncbi.nlm.nih.gov/22152021/ • Formulate unique solutions to evaluate and reconcile contradictions from experimental settings resulting in peer-reviewed manuscripts; such as, Castillo, CFG, Ling, MHT. 2014. Resistant Traits in Digital Organisms Do Not Revert Preselection Status despite Extended Deselection: Implications to Microbial Antibiotics Resistance. BioMed Research International 2014, Article ID 648389. https://pubmed.ncbi.nlm.nih.gov/24977157/
Development Plan	<ul style="list-style-type: none"> • Continue my practice.

Criteria 7: Define a problem based on a critical review of existing knowledge

- **Novice:** Can recognize a problem that is explicitly articulated or concretely given, but cannot derive one. Unaware of the depth & breadth of “the knowledge base” that is or could be relevant for the formulation of a problem. Does not recognize design features or other evidence as the basis of/support for problem articulation. Does not recognize uncertainty or how this affects the formulation of solveable problems.
- **Beginner:** Developing awareness of the depth & breadth of “the knowledge base” that is or could be relevant for the formulation of a problem. Cannot differentiate gaps in own knowledge from gaps in “the knowledge base”. Developing the ability to recognize that uncertainty may have arisen in the formulation of solutions to problems.
- **Apprentice:** Beginning to use, with guidance, the appropriate knowledge base to address a given problem. Recognizes the need to consider a wider scope of knowledge for alternative solutions to a problem common across contexts or domains. In guided critical reviews, learning to recognize that design features & evidence base are important to drawing conclusions. Recognizes the role of uncertainty in research, & that reproducibility & potential bias should be considered for every result.
- **J1 Journeyman:** Can explore & critically review the relevant knowledge base, & collaboratively articulate a problem based on that review. Reviews include assessment of relevance from (potentially) ancillary domains, bias, reproducibility, & rigor;

recognizes when appropriate & inappropriate methodology is used. Recognizes when incomplete review is provided (by themselves or by others). Can discern reproducible from nonreproducible results; can identify major sources of bias throughout the knowledge base.

- **J2 Journeyman:** Independently defines & articulates a theoretical or methodological problems based on a critical review of the relevant knowledge base(s). Knows the hallmarks of questionable research hypotheses & misalignment of testing/statistics with poorly articulated research problems; consistently finds & identifies sources of bias. Articulates when appropriate & inappropriate methodology is used/reported. Critical review & problem articulation integrate diverse disciplinary perspectives when appropriate/adaptable.

Evaluation	J2 Journeyman
Justification	<ul style="list-style-type: none"> • Published 103 peer-reviewed manuscript in conferences / journals / monographs / reference works, to date (December 1, 2020); including review articles. • Published an article on the theoretical and philosophical basis of modelling and simulation - Ling, MHT. 2016. Of (Biological) Models and Simulations. <i>MOJ Proteomics & Bioinformatics</i> 3(4): 00093.
Development Plan	<ul style="list-style-type: none"> • Continue my practice and improve my skills and knowledge in the field.

Criteria 8: Hypothesis generation

- **Novice:** When directed, follows instructions to test hypotheses; does not generate them & may not recognize them without explication. Uses the default settings of software & other tools, rather than a hypothesis, to guide any analysis. Does not question methods to be used, or assumptions of methods that are used.
- **Beginner:** When directed, uses the default settings of software, tools, or the GUI to test hypotheses in preplanned analyses; does not generate testable hypotheses. Does not recognize that hypotheses may be generated & tested within the intermediate steps of an analysis. Developing the understanding that all methods involve assumptions.
- **Apprentice:** With guidance, can: 1. leverage tools, software, data & other technologies (GUI/programming) to test hypotheses; & 2. generate hypotheses based on either the data or the technology, but not their combination/synthesis. Hypothesis generation possible in highly concrete & fully parameterized problems; developing the ability to identify whether a given hypothesis – including one of their own – is testable. Learning to recognize that experimental design & design of software/programming solutions include hypothesis generation to some extent. Developing the abilities to identify, & plan to address, assumptions that different hypotheses necessitate.
- **J1 Journeyman:** Collaboratively integrates hypothesis generation into the consideration of literature, data & analysis options. Seeks appropriate guidance in the synthesis of data & technology to generate novel, testable hypotheses. Considers the process of hypothesis generation & testing to be iterative when this is appropriate. Hypothesis generation is done with consideration of reproducibility & potential for bias, & takes into account the most clearly relevant literature; recognizes that less-obviously relevant literature may also be informative for hypothesis generation.

- **J2 Journeyman:** Independently generates testable hypotheses that are scientifically innovative as well as feasible (possible for economic reasons, time, impact, etc.). In own & others' work, recognizes that, & articulates how, hypothesis generation from planned & unplanned analyses differ in their evidentiary weight & their need for independent replication. Fully explores all relevant knowledge base(s) to support rigor & reproducibility, & to avoid bias, in the generation of hypotheses.

Evaluation	J2 Journeyman
Justification	<ul style="list-style-type: none"> • Routinely formulates and manage student projects (from high school to graduate level) given constraints. • Scientific innovation and rigors is justified by resulting peer-reviewed publications.
Development Plan	<ul style="list-style-type: none"> • Continue my practice.

Criteria 9: Experimental Design

- **Novice:** Can recognize concrete features of experiments only if they are described/given, and they match basic design elements (e.g., dependent, independent variables). Cannot design data collection or experiments. Unaware of covariates or their importance in analysis or interpretation. Does not recognize the importance of design, data collection, data quality, storage/access, analysis, & interpretation to promote rigor & reproducibility in experimental design.
- **Beginner:** Can identify salient features of experiments that are described/given if they match previously encountered design elements, but cannot derive them if they are not present. Recognizes covariates if mentioned, but does not require formal consideration (or justification) or evaluation of covariates. Does not recognize that one experiment alone cannot adequately address meaningful biological research problems. Understands that experimental design involves the identification, gathering, storing, analyzing, interpreting, & integrating of data & results.
- **Apprentice:** Can match the correct data collection design to the instruments & outcomes of interest. May include/exclude covariates, or other design features, "because that is what is done", without being able to justify their roles in the hypotheses to be tested. Developing the understanding that weak experimental design yields weak data & weak results. Needs assistance in conceptualizing covariates & their potential roles in the planned analyses. Beginning to recognize that, & can explain why, just one study is usually insufficient to answer a given research problems/solve biological problems adequately. Follows templates for the identification, gathering, storing, analyzing, interpreting & integrating of data. Learning to consider reproducibility & rigor in experimental design, & to question templates that do/do not include these concepts.
- **J1 Journeyman:** Recognizing that explicit attention to experimental design will result in more informative data; designs experiments in consultation with experts in content & statistics. These experiments may include power calculation considerations, if relevant; modeling requirements; measurement/sampling error & missing data. Collaboratively designs experiments that address the need for reproducibility & sensitivity analysis. Learning to conceptualize pilot studies & sensitivity analyses. Learning to adapt problems so that hypotheses can be generated & made testable via experiments.

- **J2 Journeyman:** Independently designs appropriate & reproducible experiments & other data collection projects, using methodologies that are aligned with the testing of specific hypotheses. Consistently identifies & justifies choices of instruments & outcomes (& covariates if relevant). Collaborates with experts as needed on appropriate use of advanced methods, including accommodating measurement & sampling error, attrition (if needed) & modelling requirements; can adapt complex problems so that hypotheses can be generated & made testable via experiments. Understands & can exploit the strengths & weaknesses of experimental design, data & modeling approaches with respect to the biological problem under consideration. Uses pilot studies & sensitivity analyses appropriately.

Evaluation	J1 Journeyman
Justification	<ul style="list-style-type: none"> • Published 103 peer-reviewed publications as of December 1, 2020.
Development Plan	<ul style="list-style-type: none"> • Establishment of more collaborations. • Update knowledge on latest instruments and experimental techniques. • Stronger emphasis on covariates.

Criteria 10: Identify data that are relevant to the problem

- **Novice:** Uses data, as directed. Does not find relevant data; cannot describe what makes data or a given data resource “relevant” to a given problem.
- **Beginner:** Correctly uses data that are provided or can follow a script/“recipe” to obtain (access, manage) relevant data to which they are guided. Cannot determine whether a given data-set or type is relevant for a given problem, but is developing an awareness that not all data are equally relevant, or equally well suited, to all research problems. Developing awareness of the features of data/data resources that constitute “relevance”, & that these features must be assessed before choosing data to use.
- **Apprentice:** Can initiate a search for data & will ask if uncertain about the relevance for any given problem. Learning how to identify, & evaluate strengths & weaknesses of, data resources, to determine whether a given data-set or -type is relevant for a given problem; &, with guidance, how to leverage these to address given research problems. Learning how reproducibility can be affected by the choice (& features) of data.
- **J1 Journeyman:** Collaboratively identifies relevant data resources. Understands the relative strengths & weaknesses of data-sets & -types for addressing their specific problem. Learning to address & formulate scientific problems (based on recognized gaps in the knowledge base) utilizing relevant data resources. In own & others’ work, recognizes that, & articulates how, choices for data (collection or use) require assumptions & justification, & must yield reproducible results.
- **J2 Journeyman:** Identifies data that are directly relevant to a problem of own or others’ devising. Consistently identifies, & evaluates strengths & weaknesses of, a variety of data resources that can address a problem or help to formulate it more clearly; recognizes if the necessary data do not yet exist. Justifies the relevance of any given data-set to a problem in terms of their individual strengths & weaknesses. Articulates hypotheses, & designs experiments, that leverage strengths in the data; includes triangulating data or results to address weaknesses in the data. Identifies whether data appropriate to the specific scientific question were used when reviewing proposals, protocols, manuscripts, &/or other documentation describing data, & research results.

Evaluation	J2 Journeyman
Justification	<ul style="list-style-type: none"> Several publications are re-using of existing and publicly available research data and models; for example, <ul style="list-style-type: none"> Wang, VCC, Kamarudin, NJ, Tan, XT, Ramesh, A, Chew, SSM, Murthy, MV, Yablochkin, NV, Mathivanan, K, Ling, MHT. 2020. A Case Study using Mitochondrial Genomes of the Order Diprotodontia (Australasian Marsupials) Suggests that Single Ortholog is Not Sufficient for Phylogeny. EC Clinical and Medical Case Reports 3(9): 93-114. Cheong, KC, Hon, RYH, Sander, CJ, Ang, IZL, Foong, JH, Ling, MHT. 2020. A Simulation Study on the Effects of Media Composition on the Growth Rate of Escherichia coli MG1655 using iAF1260 Model. Acta Scientific Microbiology 3(8): 40-44. Teng, RSY, Kwang, JCY, Chin, ASQ, Sander, CJ, Ang, IZL, Foong, JH, Cheong, KC, Hon, RYH, Ling, MHT. 2020. Correlation Analysis on Transcriptomes from Published Human Skin Studies Show Variations between Control Samples. EC Clinical and Medical Case Reports 3(6): 143-146. Ling, MHT, Poh, CL. 2014. A Predictor for Predicting Escherichia coli Transcriptome and the Effects of Gene Perturbations. BMC Bioinformatics 15: 140.
Development Plan	<ul style="list-style-type: none"> Continue my practice.

Criteria 11: Identify & use appropriate analytical methods

- Novice:** Uses methods that are provided & in a given order (i.e., a pipeline; & treats workflows; as if they are pipelines). Does not identify relevant methods; cannot describe what makes a method “relevant” to a given problem. Unaware that methods & software have default settings. Does not question propriety, assumptions, or order of methods that are employed; focus is on the superficial attributes of given methods & protocols.
- Beginner:** Uses given methods, as directed, & learning about the concepts of pipelines & workflows; still uses workflows as if they are pipelines, but beginning to attend to decision points. Learning to recognize pros & cons of methods/software, but cannot yet effectively compare, evaluate, or rank them. Becoming aware of the default settings of software or methods & their effects on results; & beginning to explore & inquire about tailored settings. Understands that more than one method/software may be available to deal with a given problem or data type, but can’t choose effectively. Learning about similarities & differences across methods, & that choices (particularly of multiple methodologies for one question) should leverage independence of methods to support reproducible results.
- Apprentice:** Can identify methods, software, & pipelines that are relevant for a given problem; seeks guidance about the best approach. Learning to evaluate/rank & justify alternative methods in terms of general features of their efficiency & relevance for the given research problem. Beginning to recognize that a “pipeline” involves only the choice of which one(s) to use; while a “workflow” requires many choices & decisions. With guidance, seeks to identify & implement appropriate workflows to address given research problems. Learning how reproducibility can be affected by the

choice & implementation of methods, including independent replication of essentially the same method vs. independent replication using diverse methods.

- **J1 Journeyman:** Collaboratively considers the knowledge base, & features of the relevant data & analysis options, in identifying the most appropriate approach(es) to tackle a scientific question. Uses appropriate analytic methods, pipelines, & workflows, recognizing, & taking advantage of the fact, that these may represent distinct approaches to the same problem. Knows when & how to control false discovery rates to promote reproducible results across methods. In own & others' work, recognizes that, & articulates how, choices for methods, pipelines, & workflows require assumptions & justification, & must yield reproducible results.
- **J2 Journeyman:** Recognizes if/when the necessary methods, pipelines, & workflows to tackle a scientific question do not yet exist. Consistently controls false discovery rates to promote reproducible results. Identifies whether appropriate analytical methods were used when reviewing proposals, protocols, manuscripts, &/or other documentation describing methods, pipelines, workflows, & research results.

Evaluation	J2 Journeyman
Justification	<ul style="list-style-type: none"> • Published 103 peer-reviewed publications as of December 1, 2020. • Many publications have their data sets publicly available for sharing and re-use (criteria 3). • Reviewer for multiple journals (see criteria 1).
Development Plan	<ul style="list-style-type: none"> • Continue my practice.

Criteria 12: Interpretation of results/output

- **Novice:** Treats the output of a program as the final/complete result – with no interpretation required - & is unaware of the concepts of validation & cross-validation or their importance for interpretation of results/output. Uses the p-value to indicate “truth” in statistical analysis. Overinterpretation is typical. Unaware of the importance of false discovery rate controls. Does not seek coherence in/recognize incoherence of their results with the analysis plan or pipeline; is unable to align methods, results, & interpretation.
- **Beginner:** Interpretation of results depends on p-values, but understanding of p-values is incomplete. Learning to recognize that interpretation of output critically depends on methods used & the pipeline in which the results are obtained. Developing awareness of false discovery rate controls. Learning that the interpretation of their immediate results could be an interim step in an overall problem-solving context.
- **Apprentice:** Seeks guidance to interpret results/output, including considerations of alignment of methods & results. Understands that the p-value represents evidence about the null hypothesis, not the study hypothesis, but does not consistently avoid reification. Recognizes that, but does not always act as if, very small p-values are not “highly significant results”. Can apply false discovery rate controls, but does so only when reminded/ required. Recognizes when the interpretation of their immediate results is an interim step in an overall problem-solving context.
- **J1 Journeyman:** Can discern, based on immediate results, methods & hypotheses, whether more experiments &/or data processing are required for robust result interpretation; collaboratively uses the appropriate knowledge base & data resources

to interpret results; resists reification & is committed to good-faith efforts to falsify hypotheses. Consistently & appropriately uses false discovery rate controls.

- **J2 Journeyman:** Interprets own & others' results critically & with respect to the analysis plan; seeks/promotes alignment of methods, results, & interpretation. Prioritizes interpretable & reproducible results above any other outcome (e.g., publication or completion of tasks/project), & insists on false discovery rate controls & other sensitivity analyses in all work. Avoids problems that can arise in the interpretation of results, including bias, reification, & other failures of positivism. Is able to evaluate the quality & appropriateness of procedures, statistical analyses, & models when reviewing papers & projects/proposals, based on the writers' – & own — interpretation of results.

Evaluation	J2 Journeyman
Justification	<ul style="list-style-type: none"> • Published 103 peer-reviewed publications as of December 1, 2020. • Reviewer for multiple journals (see criteria 1).
Development Plan	<ul style="list-style-type: none"> • Continue my practice with more emphasis on applying the philosophy of research methods.

Criteria 13: Draw & contextualize conclusions

- **Novice:** Does not draw appropriate conclusions from given results; without direction, will not even contextualize conclusions with the protocol that was followed. Not aware of the difference between conclusions about the null hypothesis & those about the research hypothesis. Conclusions may over- or under-state results & be driven by p-values or other superficial cues. Does not recognize the importance of identifying & acknowledging methodological limitations, or their implications, for conclusions. Does not or cannot apply rules of logic to scientific arguments, & commits logical fallacies when drawing conclusions.
- **Beginner:** Learning fundamentals of how appropriate conclusions are drawn from results, but may not be able to draw those conclusions from given results themselves. Learning to differentiate between conclusions about the null hypothesis & those about the research hypothesis. Learning why p-value-driven conclusions, & the lack of false discovery rate controls, are not conducive to reproducible work. Conclusions are generally aligned with given results, but when multiple methods are used, does not recognize the dependencies among methods that appear to reinforce, but actually replicate, results. Conclusions are neither fully contextualized with the rest of a document (write-up, paper, etc) or study/ experiments/paradigm (contextualization for coherence), nor with the literature (critical contextualization).
- **Apprentice:** With guidance, can draw conclusions in own work that are coherent with the research hypothesis/hypotheses & across the entire manuscript/writeup (as appropriate). Learning to critically contextualize results; is able to draw the most obvious conclusions, but struggles to see patterns, or draw more subtle conclusions. Learning that “full” contextualization of conclusions requires consideration of limitations deriving from methods & their applications, & their effects on results & conclusions. Learning to recognize how independence of multiple methods applied to similar data/problems supports reproducible conclusions.
- **J1 Journeyman:** Can extract scientific meaning from data analysis & knows the difference between statistical & biological significance. In their own & others' work, seeks competing, plausible alternative conclusions. Can judge the scientific

importance of their results, & draws conclusions accordingly. Can draw conclusions & contextualize results with respect to an entire manuscript/writeup in a given project or study, or with literature (as appropriate). Can detect when conclusions are not aligned with other aspects of the work (e.g., introduction/ background, methods &/or results, or other experiments in the project). Gives careful consideration to limitations deriving from the method & its application in a specific study. Sees patterns, & perceives more subtle conclusions than earlier-stage scientists, & collaborates to fully articulate & motivate them. Writes the Discussion & Conclusions sections, including limitations, of own articles, with collaboration.

- **J2 Journeyman:** Expertly contextualizes results & conclusions with prior literature, across experiments or studies, & within any given document (e.g., manuscript, writeup, etc.). Strives to fully contextualize conclusions in own work, & also requires this in others' work. Draws & contextualizes more subtle conclusions than at earlier stages. Can conceptualize new experiments based on the lack of robust &/or defensible conclusions in others' work. Carefully considers consistency of conclusions with the other parts of own or others' work.

Evaluation	J2 Journeyman
Justification	<ul style="list-style-type: none"> • Published 103 peer-reviewed publications as of December 1, 2020.
Development Plan	<ul style="list-style-type: none"> • Continue my practice.

Criteria 14: Communication

- **Novice:** Does not communicate scientific information clearly or consistently; is unaware of community standards for scientific communication. Generally relies on lay summaries to support own communication; does not recognize that using original literature strengthens scientific communication. Does not differentiate appropriate & inappropriate scientific communication, nor understand the ethical implications of each.
- **Beginner:** Learning both to recognize the value of clear communication, & about the role of communication in sharing & publishing research, data, code, data management, tools & resources. Developing an awareness of community standards for scientific communication, & that these include documenting code, annotating data, & adding appropriate metadata. Does not adapt communication to fit the receiver. Learning to differentiate appropriate & inappropriate scientific communication, but does not yet understand that transparency in all communication represents ethical practice, even when the desired results have not been achieved.
- **Apprentice:** Understands the roles of sharing & publishing research, data, code, data management, tools & resources in scientific communication. Seeks guidance so that own communication is coherent, accurate, & consistent with community standards (e.g., following FAIR (Findable, Accessible, Interoperable, and Reusable) principles; ensuring socially responsible science). Learning to document code, annotate data, & add appropriate metadata – & the importance of these (as appropriate given their research/context) for sharing & integration. Learning the importance of adapting communication to fit the receiver, seeking opportunities to practice this. Learning that transparency in all communication represents ethical practice, even when the desired results have not been achieved.
- **J1 Journeyman:** Consistently & proficiently uses technical language to correctly describe what was done, why, & how. Sufficient consideration given to limitations,

with explicit contextualization of results consistently included in the communication of results & their interpretation. Can adapt communication to fit the receiver; recognizes that sometimes communication must be consistent with community standards beyond their own discipline. Appropriately documents/annotates all data, code, tools, & resources for sharing, integration, & re-use. Understands that transparency in all communication represents ethical practice.

- **J2 Journeyman:** Is an expert communicator & reviewer of scientific communication; adheres to & promotes disciplinary standards for communication. Communicates in a manner that is consistent with standards across communities beyond their own discipline, as appropriate. Ensures communication is appropriate for a target audience, expertly adapting to fit the receiver(s). Communication is transparent, & appropriate to support reproducibility – & thereby, ethical practice — in every context.

Evaluation	J1 Journeyman
Justification	<ul style="list-style-type: none"> • Published 103 peer-reviewed publications as of December 1, 2020. • Many publications have their data sets publicly available for sharing and re-use (criteria 3). • Reviewer for multiple journals (see criteria 1).
Development Plan	<ul style="list-style-type: none"> • Enhance communication skills to a wider audience.

Last updated: December 1, 2020