SENIOR EXAMINER'S REPORT 2022
NATURAL SCIENCES TRIPOS, PART II GENETICS AND BBS GENETICS

Examiners: Professor Richard Durbin (Senior), Professor Cahir O'Kane (Internal), Professor Sara Goodacre (External, University of Nottingham)

1. STRUCTURE OF THE COURSE AND EXAMINATION

i) Part II Genetics (NST2GN)

The course structure for 2021-22 was the same as in 2020-21, with the same weight of individual exam components as follows:

Literature Review (8.67% of final mark) - completed during the Christmas vacation. *Research project performance and report* (16.33% of final mark) - undertaken during the Lent term. *End-of-year open book online written papers* of 24 hrs (74% of final mark) – a) Integrated paper (10% of final mark) and b) Papers 1 through 4 (each 16% of final mark) corresponding to one of four modules, respectively.

The External Examiner gave each of the 20 NST2GN students a 15-minute *viva,* in randomised order. Apart from one candidate who had tested positive for covid that morning, whose viva was by zoom, these were in person. Viva performance is not an assessment that contributes to marks, but is taken into account when considering borderline candidates.

ii) BBS

Three candidates took Genetics as their major subject (BBS course code 414): *Four written papers*, the same as Papers 1 to 4 for Part II Genetics sat alongside Part II Genetics candidates (64% of final mark), *Dissertation* (20% of final mark). *Minor subject* (16% of final mark): one student took Bioinformatics (code 128), one Vertebrate Evolution (code 130) and the other Plant Genomes and Synthetic Biology (code 147). BBS students are not invited for vivas.

One BBS student took Genetics Paper 2 (BBS course code 83) as their minor subject.

iii) Module 4 (Evolutionary Genetics & Adaptation) – *a shared paper*. Genetics Module 4 is also offered as a Lent term option (aka ZL5) to those taking NST2 Zoology and NST2 Plant Sciences, and as a major subject option to students taking Part II BBS Zoology. The conduct of the examination was identical to that observed for Genetics candidates. The scripts from Zoology students were processed by the Department of Zoology , whereas the scripts from the single student taking module 4 as an NST2 Plant Sciences paper were marked with the Department of Genetics students.

2. NUMBER OF CANDIDATES SITTING GENETICS PAPERS:

20 NST2GN candidates (not including one who withdrew before the examinations)

3 NST2BBS major (Genetics)
1 other NST2BBS students taking a Genetics minor subject

11 NST2 Zoology and NST2BBS with Zoology as major subject

Table 1: Numbers of candidates sitting Genetics Papers

Paper	NST2GN	BBS GN major	BBS minor	NST2PL	Zoology*	Total
1	20	2*	0	0	0	22
2	20	3	1	0	0	23
3	20	3	0	0	0	23
4	20	3	0	1	11	35
IP	20	0	0	0	0	20

- This does not include one BBS Genetics major student who withdrew from Paper 1 during the exam due to illness, without submitting any scripts for assessment.

All papers were taken online, with scripts being submitted through Moodle.

3. SETTING THE WRITTEN PAPERS

The form of the exam papers was identical to that in the previous year 2020-21. It was decided before the start of Michaelmas term 2021 that once again the exams would be held online and open book, but that rather than having a 24 hour window to complete each paper as in 2020-21, candidates would be given a *5 hour window* (with proportionate extensions for those with University approval) for Papers 1 to 4 in which they were asked to answer three questions out of seven, and a *3 hour window* for the Integrated Paper (IP) in which they were asked to answer one question out of seven.

This information was shared with the class in a briefing on 6 October 2021 and also posted on Moodle. Further briefings were made to the class (9 November 2021 and 8 February 2021) but the form and conduct of the examinations did not change from that presented at the start of the year.

A call for suggested questions was circulated to all lecturers on 21 January 2022. Question setters were also required to supply outline answers. Problem questions were double-checked by colleagues with the relevant academic background. The Genetics papers were drafted by the examiners in consultation with each Module Organiser. Revisions were requested from some lecturers. Those approached responded quickly and effectively. Drafts of the Genetics papers were shared with the External Examiner at the start of April, who approved the questions subject to minor changes. Rubrics were based on a template agreed across Part II NST Biology departments. All paper rubrics listed the NST2GN subject code and the relevant NST2BBS codes. In addition, Paper 4 also carried codes for NST2ZO and NST2PL (but only the Genetics paper title i.e. "Paper 4: Evolutionary Genetics & Adaptation"). The finalised Paper4/ZL5 was shared with and approved by the Part II Zoology Senior Examiner.

The Senior Examiner and the Exam Administrator were nominated as Departmental contacts for submission of examination papers.

4. CONDUCT OF ASSESSMENTS

i) Coursework:

NST2GN Literature Review and Project report. The deadline for the Literature Review was 19 January 2021 and for the Project Report 28 March 2021. NST2GN candidates were asked: (i) to submit the title and scope of their Literature Review, as agreed in a meeting with the PI hosting their research project, by the end of Michaelmas Full term; and (ii) to arrange a feedback meeting with the PI, to discuss their Project Report, by the end of the Lent Full Term. ***BBS Dissertation.*** Deadlines were as set by the Faculty of Biology: Title/proposal approved - 9 November 2021; last day for title change - 18 March 2022; submission - 29 April 2022.

All coursework was double-marked, and differences of >10% were resolved by discussion between the original assessors. All coursework was submitted to Turnitin and the output files reviewed by the senior examiner. No cases of plagiarism, or poor scholarship requiring action, were identified in the coursework by (or independent of) Turnitin. One student required a literature review deadline extension and two students required project report deadline extensions for health reasons.

ii) Written papers

The written papers took place online with papers being downloaded from and scripts uploaded to Moodle within 3-hour (for the IP paper) or 5-hour (for Papers 1 to 4) assessment periods starting at 9am on 27 (IP) and 31 (Paper 1) May and 1, 3, 6 June 2021 (Papers 2 to 4). Internal examiners and the Exam Administrator (Amy Bains) were available via email to answer any queries, but all queries were to the Exam Administrator and related to illness or issues with submissions rather than exam content. One candidate fell seriously ill during one exam and had to withdraw from that paper. In one case, where a student only submitted two scripts via Moodle when three were expected, the Exam Administrator confirmed by email that only two had been submitted. Otherwise, script submission via Moodle and use of G Suite form for coversheets proceeded smoothly.

iii) Marking and return of scripts

At the end of each exam period, scripts were downloaded from Moodle, sorted by the Senior Examiner, and allocated to assessors to mark via personal Google Drive folders with access only for the assessor and the exam administrator. This process ran smoothly, with assessors confirming that type-written scripts were much easier to mark. Almost all assessors were prompt, returning marked scripts within the prescribed 48-hour turnaround period. However, there were a few short term delays with respect to marking over weekends.

In parallel the scripts were processed through Turnitin by the Exam Administrator, with the Turnitin officer evaluating outputs. Several answers showed evidence of poor scholarship,

and one answer for one question was referred to the Chair of Examiners. This was determined to be a minor breach of the regulations and was dealt with by a marking adjustment.

The internal examiners undertook checks targeted to inexperienced Part II assessors. With the agreement of the External Examiner raw marks for two questions were moderated. The External Examiner was fully informed about the overall marking process.

iv) Shared paper (Paper4/ZL5)

Scripts from candidates who sat Paper4/ZL5 as part of NST2ZO and NST2BBS with Zoology as major subject were processed by the Department of Zoology and assessed separately. Support during the 24-hour period was offered uniformly to all candidates by Genetics in concert with Zoology. Genetics collected exam e-covers recording the questions answered by all candidates and submitted the relevant spreadsheet to the Department of Zoology, but scripts for Zoology students were processed by the Department of Zoology.

v) External Examiner and *vivas*

The External Examiner carried out *vivas* for the 20 NST2GN candidates in person on 14 June 2021. She was provided in advance with all scripts, coursework and assessment material of both the NST2GN and BBS candidates.

5. PERFORMANCE OF CANDIDATES

i) Written papers

Table 2: Question choice & average mark per question in written papers (all candidates)

Paper 1	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Overall
N	5	0	2	11	16	16	16	66
mean	66.2	-	49.0	62.7	65.5	64.6	65.1	64.3
stdev	7.3	-	8.5	9.7	6.2	8.7	6.1	6.4
Paper 2	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Overall
N	18	17	5	10	3	3	16	72
mean	65.6	65.2	62.2	65.7	61.3	67.7	59.5	63.8
stdev	11.4	9.0	11.2	9.3	5.9	5.1	15.0	8.7
Paper 3	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Overall
N	14	19	10	2	17	2	5	69
mean	67.6	64.6	67.6	65.0	68.1	56.5	63.2	66.2
stdev	4.9	14.6	7.6	0.0	7.0	2.1	8.2	6.8
Paper 4	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Overall
N	17	11	17	6	14	4	2	71
mean	57.5	65.4	61.9	60.8	64.9	67.5	75.0	61.7
stdev	7.5	5.4	11.8	11.6	4.4	6.7	14.1	9.7
IP	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Overall

N	4	0	7	1	1	2	5	20
mean	70.0	-	67.1	55.0	68.0	70.0	65.2	67.0
stdev	6.3	-	7.6	-	-	2.8	11.0	7.8

As in previous years, the number of answers to individual questions was highly variable (Table 2). Two questions did not receive any answers, and two were attempted by only one candidate. Ten questions were each attempted by ≥15 candidates.

ii) Coursework

The average mark for the Literature Review was 64.9% (SD 6.0) and for the Project Report was 67.2% (SD 9.4). For the BBS Dissertation, the average mark was 67.5% (SD 12.9).

6. THE FINAL EXAMINERS MEETING FOR PART II GENETICS was held on Wednesday 15 June 2022. The mark distribution across all papers was reviewed. All marks were awarded on a scale of 100 according to the marking scheme and class boundaries approved by the Committee of Management for the Natural Sciences Tripos. There was some discussion regarding the highest-ranked 2-i candidates: two candidates were moved into the first classification by unanimous agreement of the examiners. The 2-ii and single third class candidates were also thoroughly discussed. The External Examiner accepted the results as robust. The Thoday Prize was awarded to the best overall performance. The grade roster was uploaded to CamSIS on 16 June 2020 and the mark book and class list also sent to the Student Registry. Following this, in finalising documents, an error was identified in transcribing and accumulating a single mark for a single candidate, and this was communicated to the Student Registry, the student affected and other students whose ranking changed; no changes in degree class resulted from this correction.

7. OUTCOME OF THE PART II GENETICS EXAMINATION

i) NST2GN class distribution

First class: 25% (n = 5)
Upper Second class: 60% (n = 12)
Lower Second class: 10% (n = 2)
Third class: 5% (n = 1)

ii) BBS (Genetics major subject)

The completed mark book was submitted to the Faculty Office on 24 June.

8. STATISTICS FOR THE EXAM RESULTS

i) Class composition

20 NST2GN candidates: 16 NST (10 female, 6 male), 4 MVST (2 female, 2 male).

3 BBS Genetics major candidates: 2 NST (both male), 1 MVST (male); plus 1 BBS NST student (male) taking a single Genetics paper as their minor subject.

13 Zoology/Plant Sciences candidates taking shared Paper 4/ZL5: 11 NST2ZO, 1 Zoology BBS, 1 NST2PL.

These numbers do not include one NST2GN student who intermitted during the year and did not take exams, but they do include an NST2GN student who had intermitted during 2020-21 and returned during 2021-22 and took exams to complete their degree.

ii) Distribution of classes and movement between classes (NST2GN only)

No information available as the Part IB class was subject to formative assessment in 2019-20.

iii) Mean marks in each exam component by gender (NST2GN only)

	Female		Male	
N=	12		8	
	Mean	SD	Mean	SD
Literature Review	65.8	5.3	63.6	7.1
Project	70.0	10.3	63.0	6.2
Examinations	66.7	6.1	63.0	5.5
Total	67.2	6.1	63.0	5.0

iv) Mean marks in each exam component by background (NST2GN only)

	NST		MVST	
N=	16		4	
	Mean	SD	Mean	SD
Literature Review	64.4	6.4	66.9	4.1
Project	66.3	10.0	70.9	5.4
Examinations	65.1	6.2	65.6	5.8
Total	65.2	6.2	66.6	5.2

v) Representation in classes by gender (NST2GN only)

	Female		Male	
N=	12		8	
	Count	Percent	Count	Percent
1	4	25%	1	13%
2i	7	44%	5	63%
2ii	0	0%	2	25%
3	1	6%	0	0%

vi) Representation in classes by background (NST2GN only)

	NST		MVST	
N=	16		4	
	Count	Percent	Count	Percent
1	4	25%	1	25%
2i	10	63%	2	50%
2ii	1	6%	1	25%
3	1	6%	0	0%

9. REPRESENTATIONS & APPEALS

No formal representations were received. The Senior Examiner responded to four information-only requests for various details including verification of marks and general information about marking as in Appendix A. Information was provided to support two applications for an allowance.

10. RECOMMENDATIONS

While the online open-book examinations were generally felt to work well, in particular with the revised shorter (5 hour) time window, there was some concern about the Integrated Paper, for which the questions typically address broad topics for which reviews or perspective pieces are available online. Following discussion between the examiners and with other members of the department, it is recommended to revert this paper to be taken closed book, still in a 3 hour window.

11. ACKNOWLEDGEMENTS

I would like to thank my colleagues, Prof. Cahir O'Kane and Prof. Sara Goodacre, and the Exam Administrator, Amy Bains, for their support and hard work in connection with the exams. Special thanks are due to Prof. Marisa Segal (Senior Examiner in 2020-21) for her advice and passing on a clear structure and documentation for carrying out the process, and to Dr. David Summers for his invaluable wisdom and experience, and for acting as Turnitin officer. Finally, I would also like to thank all participating assessors and project supervisors.

Professor Richard Durbin (Senior Examiner) 2022

APPENDIX A: SENIOR EXAMINER'S REPORT 2022, Part II GENETICS AND BBS GENETICS, NATURAL SCIENCES TRIPOS

IP: Integrated Paper
Paper 1: Genomes, Chromosomes and the Cell Cycle
Paper2: Human Genetics, Genomics & Systems Biology Paper 3: Developmental Genetics
Paper 4: Evolutionary Genetics & Adaptation

Paper	Number	Num answers	Assessor	Summary of expectations for the question	How the answers fulfilled expectations
IP	1	4	David Summers	This topic can be approached in a variety of ways. Points that might be considered include: • A viral pandemic is an acute problem that can overwhelm healthcare systems with alarming rapidity while the antibiotic "crisis" is more long-term. • A viral pandemic has the potential to collapse national and international economies. It is harder to imagine how antibiotic resistant bacteria would have a similar effect. • Vaccination strategies have a proven track record in countering a viral pandemic. Vaccine design is faster and more flexible than the design of conventional antibiotics. Vaccination against bacterial disease has considerable untapped potential; use against tuberculosis might be considered here. • Effective antibiotics are essential in routine surgery (e.g. hip replacement) and immune-supressing treatments (e.g. cancer chemotherapy). A loss of effective antibiotics would have widespread and long-lasting effects in healthcare. • It is unrealistic to imagine an antibiotic to which resistance cannot evolve. Equally there can always be a new viral variant. • The floating genome is a highly developed mechanism for the rapid dissemination of antibiotic resistance. • The threats from both are serious but distinct in their natures and consequences.	Weaker answers were unbalanced, focussed mainly on antibiotic resistance with only occasional reference to viral pandemics. Better answers were well-balanced and compared the two threats in a range of contexts (e.g., scientific, medical, economic, social).
IP	2	0	Erik Clark	No answers received,	No answers received.
IP	3	7	Ben Steventon	I expected answers to this question to provide two main components: 1) A broad set of examples where genetic engineering technologies have been applied in medicine, agriculture and/or medicine. Each example should contain evidences of potential for positive and negative impacts on society. 2) A critical discussion of how and when the public should be consulted on the ethical implications of the examples given.	1) Overall, this was addressed well in the answers and proved a good question to draw out general knowledge and understanding of the underlying genetics of each application. 2) Answers to this were mixed and there were few examples of critical thinking in this regard.
IP	4	1	John Welch	The question was an invitation to consider the use of simplified models in genetics (where "models" could be mathematical, statistical, verbal etc.). The distinction between biology and physics could have been discussed. Examples could have been given, stating each model's simplifying assumptions (i.e. how they are not accurate descriptions of the world), how, exactly, the model is useful to biologists (useful with respect to either understanding or	The question received only one answer, which was a summary of parts of two M4 lectures on the evolutionary maintenance of sex. Various hypotheses for the evolutionary maintenance of sex were discussed and each judged unsatisfactory in some way, as in the lectures. No broader points were raised.

				application), and ideally, how the usefulness is connected to the simplifying assumptions.	
IP	5	1	Cahir O'Kane	A good answer would review the rationale behind the use of model organisms, and critically review many of the uses to which they have been put, with pertinent examples, and their advantages and limitations. Issues for discussion include: the meaning of "model" and the extent to which other organisms (across the living world, not just humans) can be modelled; technical and ethical advantages; limitations; pros and cons of different models; comparisons to cell-based systems including stem cells/iPSCs; uses for mutagenesis and gene discovery, maybe discussion of unbiased vs reverse genetics. A good answer would show critical awareness of the degrees to which features can be modelled, and to which models can be extrapolated.	One answer was received. This showed a good breadth of knowledge of the above issues, and good critical awareness across a wide range of them. However it was quite human-centred, and could have done better on comparisons of models, their roles in gene function discovery and whether this is still of value, and the extent to which features can be modelled or extrapolated.
IP	6	2	Aylwyn Scally	The answers submitted were thoughtful and well argued, and generally similar in their approach and conclusions. In each case they defined genetic ancestry carefully, albeit in somewhat theoretical terms rather than how it would actually manifest in a specific clinical context. They gave some nice examples of disease-ancestry associations and discussed issues with ancestry inference. Differences between genetic ancestry, race and self-ID ethnicity were also discussed (although interestingly neither used the latter term). However both essays took the terms of the question itself at face value; in a very strong answer I was looking for these to be challenged and assessed. Is it clear what genetic ancestry means in different contexts - for example, are the forms of genetic ancestry defined or described the same as those used implictly in GWAS studies? Has the boat sailed and is genetic ancestry already implicitly and important consideration? There were also one or two important points that neither really addressed head on. For example, the underlying question of whether or how might genetic ancestry provide risk info over and above knowledge of risk alleles from GWAS. To what extent do PRS estimates include genetic ancestry, for example (depending on environment, cohort studied etc)? The issue of distinguishing genetic ancestry-associated risk from environmental risk in GWAS setting, and quantifying these relative effects.	see left for integrated feedback
IP	7	5	Richard Durbin	Relevant points include: • The reference genome provides an (essentially) complete list of parts, introducing and enabling genomic science. It supports annotation to organise genetic knowledge the mapping of experimental sequence data (RNA-seq, ChIP-seq, Hi-C, ATAC-seq etc.) which allows experimental results to be connected to each other and genes and gene regions. • Sequencing for genetic variation helps get at the genetic component of phenotypic variation	The best answers made many of these points, the least good just reproduced parts of lectures addressing some of the words in the question. Marking was neutral as to which was taken to be more important, although all but one answer argued for the next million genomes being more important than the reference – perhaps the students, being geneticists, like the study of variation, and the importance of the reference is now taken for granted. The other answer argued that the reference is more important for

				and disease. One million genomes is about where we are now with UK BioBank and recent GWAS. Foundation for genomic prediction and mendelian randomisation, which is important for drug development (but so is the knowledge from the reference). There is not enough natural variation in humans to fully understand gene function, or e.g. find knockout phenotypes of all genes.	scientists, and the variation for societal impact, which was a nice way to think about it.
P1	1	5	Marisa Segal	The concept of model systems in research. Critical review of the impact of the cdc screen and how it shaped our understanding of fundamentals in cell cycle control and their validity to all eukaryotic organisms. Molecular insight and critical control points. In turn, how those advances have shaped our view of cancer as a disease of the cell cycle and bring us to present research, novel therapies, systems view, etc.	All answers were on target but varied in how the showcased key concepts with some distance (or not) from lecture material.
P1	2	0	Marisa Segal	A discussion of the virtues of single cell analysis applied to cell cycle studies and, in particular, for dissecting the restriction point. Biosensors, FUCCI, transcriptional reporters. Significance of thresholds, feedback loops etc. Single cell vs. population analysis. Studies for positioning or testing the integrity of the restriction point.	No answers received.

P1	3	2	Marco Geymonat	For (a) I was expecting an essay centered on the importance of the SPB acting as a hub for the activation of the Mitotic Exit Network. In particular, the Kin4-dependent regulation of the affinity of the two-component GAP Bub2/Bfa1 for the SPB being the most upstream step controlling the activity of Tem1 (hence the entire MEN cascade). At the bottom of the cascade, the affinity of the kinase Dbf2 for the SPB is also regulated (by the kinase Cdc15) so that the protein can leave the SPB and enter the nucleolus to phosphorylate Net1 and contribute to the release of Cdc14.
				
				For (b), I was expecting to propose to somehow alter the ability of those proteins to leave the SPB (by forcing them to stick to the SPB or by impairing their ability to interact with the SPB) and check the consequences on the cell cycle progression and/or the spindle position checkpoint.
					This question was answered quite poorly: the importance of the change of affinity of Bfa1 and Dbf2 for the SPB was only slightly mentioned and the experimental approach did not really address the question.

P1	4	11	Christine Farr	Explain CENP-A role as epigenetic mark in CEN identification in many eukaryotes inc. human, in mitosis and meiosis. Explain role of centromere and importance of restricting formation. Consider concentration of CENP-A nucleosomes at the CEN and along chromosome arms. Consider evidence that ectopic human CENP-A does not usually lead to kinetochore activity. Review the multiple mechanisms limiting misincorporation and preventing neocentromeres, inc.: - maintenance through CENPA loading outside S - pathways for removal of ectopic CENP-A - establishment of multiple intermolecular CCAN contacts - importance of CEN satellite DNA - detrimental effects of spontaneous CEN formation	Most answers were on target. While first class answers considered all the points listed, in some depth, weaker answers omitted key aspects and lacked detail.

P1	5	16	Christine Farr	Explain what TADs are and how identified and validated. Consider conservation and differences across eukaryotes. Consider whether they arise actively (e.g. loop extrusion) or passively (self organisation e.g. LLPS). Appear to be hardwired, fundamental features of genomes. TAD positions correlate with many genome features suggesting functionality e.g. histone PTMs and chromatin type, replication timing, and distribution within them of genes. Consider their relationship to other features of the genome, such as A/B chromatin compartments & LADs. In mammalian genomes, regulatory elements (enhancers) often lie within the same TAD as their target genes. This has led to the view that TADs act as zones enhancer influence. Discuss fully the experimental evidence into whether, or not, TADs regulate genome function. Consider experimental disruption of TADs and TAD organisation by various methods. Illustrate points with examples. Discuss caveats/ limitations in our understanding and consider alternative possibilities.	Generally decent answers, although only a couple were really strong. Weaker answers had only a brief consideration of TADs, which did not fully explore the experimental evidence, and/or had extensive discussion of other genome features, which they failed to relate directly to TADs and to the question. Description of Hi-C method was often poor, missing key steps, or in the form of a wordy description, rather than through the use of a good diagram.
P1	6	16	David Summers	The following ideas might be explored in this essay: • The floating genome (FG) is the collection of elements, chromosomal and extra-chromosomal, that are capable of moving locations within or between bacterial cells. This includes phage, plasmids, transposons, conjugative transposons etc. • Some members of the FG are hard to define as they display the properties of more than one class. • The evolution of FG components has been described as a reshuffling of functional modules giving rise to a continuum of structures and mechanisms • Interactions within the FG enable the shuffling of genes within and between replicons. • Movement between static and mobile elements (conjugative transposons and conjugative/mobile plasmids) means that any gene can become mobile both within and outside the boundaries of bacterial species. • All of this FG activity generates genetic variation. However, this does not lead to change/evolution in the absence of selective pressure. The example of antibiotic resistance might be used to illustrate the result when both are present.	The best answers started by describing the range of elements within the floating genome and how they evolve by module shuffling. This was followed by a clear focus on the way in which interactions among elements generate the variation on which selection can act. Antibiotic resistance is a good (but not the only) illustration of this process. Weaker answers dwelt on the individual elements, saying little about their evolution or interactions.
P1	7	16	David Summers	Summary answer: • ColE1 dimers form via homologous recombination. Each dimer has two origins and so replicates twice as often as a monomer. Dimers therefore accumulate rapidly and clonally. They have a copy number approx half that of monomers and so reduce plasmid stability. • Dimers are resolved by (relatively slow) Xer-cer site-specific recombination. • Indole production (stimulated via Rcd synthesised from plasmid dimers) halts cell division and plasmid replication while the resolution takes place. • Empirical evidence suggests that neither P1 nor pT181 suffers from the dimer catastrophe. • P1 dimers, if they form, would self-handcuff very rapidly. Thus, rather than proliferating,	Most candidates were able to give at least a reasonable description of the dimer catastrophe and its antidote (Her-cer recombination plus the Rcd checkpoint). Discussions about the fate of dimers of P1 and pT181 were much more variable. This was intended to be a challenging part of the question and marks were given for intelligent discussion rather than the correct answers as determined by experiment (see above).

				they would be out-replicated by monomers. Therefore, no dimer catastrophe. • A pT181 dimer would not proliferate. It would replicate to give monomers because the Rep protein would cleave off the monomer-length displaced single strand when it reached the origin sequence at six o'clock. Therefore, no dimer catastrophe.	
P2	1	18	Gos Micklem	Opportunities over many sequencing application areas: - cheap enough to sequence everyone even in middle-income countries - this raises health insurance issues and, in authoritarian countries, privacy risks. - with suitable consent, potential for whole-population genetic studies - potential for improved disease diagnosis/ prognosis - facilitates precision medicine, - potentially more efficient healthcare - speed means that point-of-care sequencing is realistic, e.g. to identify infections and drug resistance. - point of need screening for e.g. viral infections becomes possible - facilitate many aspects of modern biology: - greater power for metagenomics/ characterisation of genetic diversity - genome sequencing and expression-based annotation - it would enable larger-scale and/or higher depth single-cell gene expression studies. - epigenetic studies: - DNA methylation - chromatin mark through ChIP-seq - deeper iCGC type studies: greater ability to order events in carcinogenesis - more comprehensive ENCODE-style studies and the ability to apply this approach to more organisms - enable more powerful characterisation of forward genetic screens If they were not already limiting factors: - sample collection (e.g. indigenous rights over plants) - sample processing/storage - possibly patient consent for widespread downstream use - perhaps the degree to which samples can be multiplexed - data storage and analysis: costs and expertise	There was a tendency for candidates to assume the question was about human genome sequencing rather than sequencing in general. They tended not to consider the limitations as they went along, often resulting in an incomplete account of limitations at the end. Almost no candidates considered the possibilities of point-of-care use when sequencing is 1000x faster, nor the fact that sample collection and preparation were likely to be limiting factors.
P2	2	17	Michael Imbeault		As in previous years this one is somewhat of a normal distribution - most answers are good but not great, and a few are excellent and a few miss the point completely. Overall satisfied with the average answer which shows that they understood both the question and the material of the course, and could think of how to apply it to a new problem.

P2	3	5	Richard Durbin	Sequence assembly works by identifying overlaps between adjacent reads. The main problem is repeats. Length is important to span repeats so as to ensure that the overlap is unique in the genome. Accuracy is important so as to be confident that the overlap is correct, not between similar but different sequences. There is a trade-off because short read methods such as Illumina are cheap and high throughput and accurate, but only ~100bp read length, so can not span repeats or phase haplotypes so result in short contigs. Long reads, from single molecule platforms such as PacBio or ONT, can be 10s to 100s kb long, but generally are high(er) error and more expensive. These can give long contiguity assemblies but need higher coverage to be (reasonably) accurate. Other long range info from short reads such as mate pairs, HiC, linked reads, genetic map markers can help scaffold shorter contigs from short reads. Recent HiFi (CCS) is both very accurate and (fairly) long, but more expensive. Combining methods, as in the recent (2022) human T2T complete assembly can give the best of all worlds. When looking for variants by assembling against a reference, accurate (e.g. Illumina) reads are better for small variants such as SNPs and short indels, long reads are better for structural variants.	Most answers identified the main trade-off, but only one showed a deeper understanding of the issues and considered related relevant issues.
P2	4	10	Richard Durbin	Diagnosis of the causative mutation for a candidate monogenic disease patient requires use of: • Inheritance pattern: family phenotypes and where possible genotypes. Discuss at least recessive, dominant, de novo + potentially others. • Phenotype – can identify candidate gene or set of genes • Sequencing data: can be exome, whole genome, or arrayCGH for CNVs – ignore common (>1% or even 0.1%) population variants as not plausible except in trans to a rare variant in a recessive. De novo variants not in parents in known gene are strong candidates. Not genotyping array/GWAS/imputation. • Previous data associating a gene to a disease phenotype: e.g. OMIM and/or ClinVar. Very important for standard diagnosis. Discovery of variant in gene not previously associated has a much higher bar. Other relevant info includes use of ACMG guidelines. Function in cells/model organisms is more relevant to identifying new genes than diagnosing individual patients. Reasons for most identified monogenic disease variants being in coding sequence are: • Coding variants are more likely to be strong effect, disabling proteins and hence gene function enough to cause single-variant disease – mutations in non-coding functional sequence are more likely to be weak effect, potentially contributing additively to polygenic disease. • Ascertainment bias: we can interpret coding variants better and aggregate across genes, and also we have mostly looked in coding sequence (though attempts have been made to scale and aggregate and look at non-coding variation for monogenic disease, with thin results, and just a few exceptions).	Most answers to part 1 were reasonable, though many included additional irrelevant material from lectures. Least well covered was the importance of previous observations of the same variant or type of mutation in the same gene. Quite a few discussed GWAS (not relevant), and model organism experiments (unlikely for diagnosis of a patient). A couple answered a different question: how to identify an unknown disease gene, rather than find the causative mutation in a patient (but still got marks because there is a big overlap). Responses to part 2 were more variable, with fewer identifying the importance of effect size and a couple of very confused/wrong suggestions.

P2	5	3	Felix Day	The best answers to this question would have described both the results and the methods used in three core case studies, with additional details coming from a wider understanding of the literature and setting the results into a broader context. Candidates could have used the studies of menarche and menopause to explain the logic of genome-wide association studies, and crucially the information that had come from these sorts of studies. In particular, when discussing the menarche, the candidates could have talked about the overlap of the loci discovered with the loci for BMI, this was not a point made by the candidates in the main, other than as a hinderance for doing Mendelian Randomisation using puberty signals. The fact that this suggests important shared pathways between adiposity and puberty was often overlooked. As were the tissues – those in the brain - that were identified as key for puberty timing. Candidates were better at explaining some of the biology revealed by the studies of menopause, and the majority covered the importance of DNA repair in the maintenance of the oocyte pool. When discussing the study of Japanese women, candidates did cover the fact that different loci had been discovered, but often missed that this was due to different genetic architecture (though some did mention the impact of drift as explaining the difference in genetic background). In general, when discussing the GWAS discovery studies candidates were stronger when talking about the methods used in the papers then they were writing about the results or interpretation of those results. In most cases the sections on the Mendelian randomisation work were better explained – with answers covering both the methodology and the results that had been seen. Only the best answers put this into some sort of public health context. The role of parent-of-origin effects on puberty timing was overlooked by many of the candidates, which would have provided a nice way to link this answer to other parts of the course through the mechanism of imprinting.	Integrated response at left.
P2	6	3	Steve Russell	Integrating data collection with computational modelling & theory. - Aims for a comprehensive & integrative understanding of biological functions and organisation - Ultimately, want to know how life works - not just the nitty-gritty of individual molecular functions but a deep understanding of the logic that underlies the molecular and cellular pathways that enable life. - The simplest view of Systems Biology is represented by the so called virtuous cycle where Data (preferably quantitative but not necessarily) is used to develop a PREDICTIVE MODEL. The model suggests hypotheses that are tested by further experimentation & data collection. These data are used to refine the model and so on. - Types of modelling approaches (Boolean, Deterministic and Stochastic) and deductive versus inductive. - Could include Synthetic biology	Only 3 essays, two were very good in terms of exemplifying systems approaches and covering the rationale and types of modelling employed. Both included examples I did not teach, showing a good understanding of how systems approaches can be useful and good use of the open book approach. One was less impressive, a short essay relying entirely on the lecture material and missing some key concepts

				- Examples could include any area where modelling approaches have increased our understanding of biological processes or the logic of regulatory circuits. Key examples in lectures were phage lambda and the repressilator.	
P2	7	16	Hansong Ma	I expected students to answer the question by summarising mtDNA inheritance and heteroplasmy transmission first, and then address how nuclear-encoded proteins ensure the uniparental inheritance, and regulate/influence mtDNA mutation levels in heteroplasmic organisms. A brief discussion of mito-nuclear interaction is also favoured.	Most of the students managed to describe mtDNA transmission patterns, and discuss at least one aspect of nuclear influence on mtDNA inheritance (e.g. uniparental inheritance). A few students addressed all the points. Overall, the answers were not too derived from what I expected.
P3	1	14	Daniel St Johnston	The topic of epistasis was covered in detail in the lectures, as well as the different interpretation of epistatic interactions between mutants in enzymatic versus developmental pathways. There were no really weak answers and all of the students seem to have understood the basic principles, although they expressed them more or less clearly. They also described most of the epistatic interactions in the dorsoventral pathway in Drosophila correctly. The better answers accurately described the use of mRNAs encoding gain of function serine protease variants to order the mutants upstream of Toll, although several quite worryingly described injecting the mRNA into the extracellular space around the embryo. The best answers included epistasis between gurken and dorsal mutants and between transplanted peri-vitelline fluid from Toll null mutant embryos to pipe mutant recipients.	This was a pretty straightforward question that tested the students' understanding of a part of the lecture course that they often find confusing, so I was pleased that all of them grasped the key points. The spread of marks is quite narrow, as there was less scope to really shine and nobody messed the question up badly.
P3	2	19	Julie Ahringer	The marks were awarded in two parts. Half were awarded for describing what P granules are (20 pts) and what is known about the mechanism of their localisation (30 pts); the other half were awarded for describing an approach to identify new genes involved in formation and localisation of P granules (20 pts), and discussion of how the functions of these new genes would be studied (20 pts), and how their relationships with known P granule regulators would be investigated (10 pts). Answers should have included that P granules are ribonucleoprotein condensates that are germline associated and important for germline fate and function. Some points were reserved for relevant information not given in the lecture. The mechanism part should have explained that localisation occurs by regulating dissolution/condensation, described how PAR-1 regulates MEX-5 by phosphorylation and that phosphorylation increases MEX-5 diffusivity, and the RNA competition model involving MEG-3. Again, some points were reserved for relevant information not given in the lecture. The experimental approaches part needed to	Overall, the question was answered well, with 15/19 answers being marked a 1st or 2.1. The first part was better answered.

				describe an approach for identifying genes including the assay (e.g, forward or reverse genetic screen, assaying P-granules using a strain that carries a fluorescent P-granule reporter). The last part, describing how the new genes would be studied, was the most poorly answered. Many general approaches were acceptable (genetic, biochemical, phenotypic, literature based). Investigating relationships with known factors should have included use of par-1, mex-5, P-granule components mutants (interdependencies) and/or proteins (interaction analyses).	
P3	3	10	Eric Miska	See right.	I am very pleased with the answers overall. There was a good spread in quality. It was clear that students had to select specific parts of my lectures to assemble a complete answer. Several did. In addition, there was some creative additions from other lectures, the reading list and other sources. Well done.
P3	4	2	Ben Steventon	See right.	There were only two answers, however both were similar in that they encouraged the inclusion of information across the module. The question was answered well, with the main points covered in both essays.
P3	5	17	Felipe Karam Teixeira	Expectations: - miRNA biogenesis relies on the transcription of structured hairpin RNAs (pri-miRNA) and depends on the activity of Dicers and that small RNAs are eventually bound by Argonaute proteins in a RISC complex, while piRNAs are generated from transcripts derived from piRNA clusters in a dicer-independent manner (a good essay will describe the ping-pong amplification loop). - miRNAs exert silence by either inhibition of translation or mRNA destabilization (both in the cytoplasm), while piRNAs can lead to the cleavage of the target mRNA in the cytoplasm (post-transcriptional gene silencing) or transcriptional gene silencing (chromatin modifications at target loci). - miRNAs control gene expression, while piRNAs mainly target transposable elements. - Restriction of piRNA pathway to the germline vs developmental role for miRNAs (a good essay will describe some of the developmental phenotypes associated with mutations in the miRNA pathway and the sterility associated with mutations in the piRNA pathway)	Most essays did a good job describing the mechanisms involved in miRNA and piRNA biogenesis and action, and contrasting/comparing the two pathways in terms of function.
P3	6	2	Marta Shahbazi	See right.	Only two students chose question 6 and I was very disappointed with both essays. They were both incomplete (half of the total length) and presented conceptual mistakes. I have been marking essays for the past 5 years and this time I have given the lowest marks.

P3	7	5	Erik Clark	The intention of this question was for candidates to demonstrate and apply their understanding of PI, R-D and TST. First, they should use simple diagrams to show how each mechanism could be used to set up a repeating array of stripes. For PI this could be achieved either by using many different concentration thresholds to read out a gradient directly, or by using hierarchical readout as in the Drosophila segmentation cascade. For R-D this could be achieved if the lengthscale of the standing waves was small with respect to the tissue length. For TST this could be achieved by a clock and wavefront. Next, they should give an example of each mechanism being used in an embryo (eg Drosophila segmentation cascade for PI, digit patterning for R-D, vertebrate somitogenesis for TST). Simply stating an appropriate example would be sufficient, but this could also be an opportunity to demonstrate outside reading or incorporate information from other lectures by providing some genetic/molecular detail. Finally, the candidates should think about the properties of each patterning mechanism and explicitly compare them - why might one perform better or worse than the others for metrics of interest (robustness, scalability, speed, flexibility, genetic complexity, etc)?	The answers showed that candidates generally had a good basic understanding of PI, R-D and TST, but were disinclined to engage with the question as written. For example, all but one candidate ignored the stipulation of establishing a *periodic* pattern, and instead copied the diagram from the lecture notes showing the patterning of a French flag pattern. No candidate compared the three mechanisms to each other explicitly, though some strengths and weaknesses were mentioned in passing. Some answers were clearly partial / unfinished. Embryonic examples, where provided, were usually appropriate, and in several cases demonstrated evidence of outside reading or incorporated material from other lectures. Rather than comparing the mechanisms to each other, answers tended to assess whether Wolpert's PI hypothesis was a good model for morphogen-mediated patterning in real embryos; fine, but not the question asked.
P4	1	17	Chris Jiggins	I was hoping for a focus on the role of low recombination in leading to high between species differentiation due to multiple factors - increased linked selection due to both background selection and selective sweeps, reduced admixture between species due to greater removal of linked alleles to introgressed alleles, and finally enhanced accumulation of divergence due to linkage of coadapted allleles. A good answer should address all these points. This was NOT a question about supergenes associated with polymorphism within species (unless a student could explicitly find an example to link this with between species divergence, which nobody did).	In general therefore this was a very poorly answered question - lots of students discussed supergenes and gave examples of polymorphism such as self incompatibility or mating system polymorphism - not related to speciation. There was only one answer that was reasonably complete. Also, there was a great paucity of any evidence for extra reading beyond the lecture notes. Overall one of the weakest sets of answers that I have seen for a long time.
P4	2	11	Nick Mundy	This question required integration of models of gene duplication with population genetics, phylogenetics and gene function analyses. The main models of gene duplication should be described and then tests applied to the situation in the question. These could include tests of selection (population genetic, phylogenetic and ecological) and functional analyses of genes.	Most answers contained good descriptions of gene duplication models but there was wide variation in how these would be tested in the context of the scenario in the question. The best answers nicely integrated population genetic methods. Analysis of gene expression patterns was rarely mentioned.
P4	3	17	Emilia Santos	I did a whole lecture on a meta analysis of the genetic basis of vertebrate pigmentation analysing the relative contribution of cis- and coding mutations to vertebrate pigmentation evolution.	I did not think my question was particularly hard and thus was expecting better responses. Over all I think that most essays were rushed and not well written often out of focus. I think some of the students might have misread the question and thought that this was perhaps a question from Nick Mundy, as the content of the responses covered the material and examples that he covered (e.g. extensive discussion of mc1r and agouti case studies instead of focusing on the broader picture). This was disappointing.

P4	4	6	Aylwyn Scally	Calculations: Overall these were mixed. Several candidates missed the factor of two due to the comparison across to lineages since divergence. Related to this, several answers mistakenly assumed the chance of an identical site in the alignment was P(t)^2 rather than P(2t). To some the degree the wording of the question was difficult because it said 'unchanged' rather than 'be identical', but this error should not have been made given a genuine understanding of the calculation and model. Nevertheless in marking (d) I did not double-penalise if the approach was correct given the answer in (c). Discussion in (b): The discussions tended to be a bit limited. All mentioned the issues of ancestral polymorphism (or equivalently, ancestral coalescent time), and of multiple mutations, but both of these were prompted in the question. There was no mention of mtDNA having a small Ne (which potentially minimises the former issue). Only one answer alluded to potential impacts of selection, and a couple mentioned potential changes in parameters such as mutation rate and generation time.	Integrated response at left.
P4	5	14	John Welch	The question was asking for a clear explanation of why the evolutionary maintenance of sex is considered a major problem in evolutionary genetics, and why Weismann's hypothesis, which held sway for so long, came to seem unsatisfactory. The question also asked for a discussion of lines of evidence that might count against this hypothesis.	This was quite a difficult question, but as in previous years, most candidates decided to summarize the lectures, in the same order, using the same examples, and reaching the same conclusion. Few if any candidates engaged thoughtfully with the part of the question about types of evidence. The major difference between candidates was how much of the lectures they chose to summarize (some discussing only the Lottery Model, some only Clade Selection, some all of the hypotheses in lecture order). Most of the material recalled was accurate, but overall, there was little to separate the best and worst answers.
P4	6	4	Bill Amos		I was generally rather disappointed with the answers, which tended to rely on listing examples more than creating a narrative that actually addressed the question set. The standard of English continues to decline. One answer was barely more than a series of note / statements. This really does not, in my view, meet the requirement to discuss.
P4	7	2	Frank Jiggins	The answer required a summary of the ways in which genetic variation could constrain the response to selection, including selection on multiple traits/genetic correlations and processes reducing variation. I was then looking for a specific system and research program to address these questions.	The strongest answers engaged fully with this second part, with a specific system and applying concepts from the lectures to a new situation. Less strong answers largely focussed on the examples and case studies in the lectures.