

MMBB 488/ 588: Genetic Engineering

Organization and goals of this course

This course is intended for students who have completed one semester of genetics, or one semester of microbiology and biochemistry. Genetic engineering in one form or another is an integral part of nearly all experimentally-based research in biology today. This course will cover the building blocks upon which current work is based. During the semester, we will discuss basic topics such as gene regulation, codon usage, and DNA replication as each pertains either to gene modification, or gene isolation. Special attention will be given to the use of gene knock-outs and replacements in both prokaryotes and eukaryotes to further our understanding of Downs Syndrome, aging, cystic fibrosis, and disease-resistance in plants. The last few weeks will detail some of the unexpected problems that been revealed by studies of transgenic organisms. We will pay special attention to how studies of unintended gene silencing have led to the development of RNAi techniques.

This course is intended to prepare you for future careers in science, medicine, and in allied fields. If there are specific topics that you feel will prepare you for your chosen future, or complete what you have learned in other courses, I will be glad to modify the lectures to teach those. Offer your suggestions to me at any time.

At least one lecture will be devoted towards discussing different approaches to reading, analyzing, and assessing scientific papers. Most of the lecture will come from research material published in various journals. If you are interested in going to the original articles, please contact me.

The following pages include a course outline, and a statement of course requirements and grading procedures. It will help if you read all of this during the first few weeks of class. Reading things like this together with the text helps build your vocabulary.

Disability Support Services Reasonable Accommodations Statement:

Reasonable accommodations are available for students who have documented temporary or permanent disabilities. Please notify your instructor(s) during the first week of class regarding accommodation(s) needed for the course. All accommodations must be approved through Disability Support Services located in the Idaho Commons Building, Room 306. 885-6307

email at dss@uidaho.edu

website at <www.access.uidaho.edu> or www.webs.uidaho.edu/taap

Please note. I am required by the university to inform you of the following policy:

1. Unexcused absence from exams or failure to turn in an assignment at the specified time will be dealt with according to University of Idaho General Catalog section M of General requirements and Academic Procedures.

2. Acts of cheating or plagiarism in MMBB 488 can result in an automatic 0 pts for that exam or assignment. A zero on any exam or assignment will result in an automatic F as a final grade in the course. All parties involved in cheating or plagiarism are responsible and may be treated equally. Cheating is defined as acquisition of answers to test questions or assigned materials in a dishonest fashion. Plagiarism is defined as 1) using or substituting another student's writing in place of your own and/or using published material without providing corresponding citations. Plagiarism includes copying or paraphrasing another person's writing without significant changes in wording. You will be presented with information in three forms. The first comes from lectures, the second from text and handout readings, and the third from the primary literature. The book is meant to provide you with background, the handouts with basic principles and convenient reference tables, and the papers with the heart and soul of the field. The lectures are distilled from the papers. Each source may complement the others, but will not substitute for the others. You should attempt to pursue each source of information and integrate it with the rest.

University of Idaho Classroom Learning Civility Clause

In any environment in which people gather to learn, it is essential that all members feel as free and safe as possible in their participation. To this end, it is expected that everyone in this course will be treated with mutual respect and civility, with an understanding that all of us (students, instructors, professors, guests, and teaching assistants) will be respectful and civil to one another in discussion, in action, in teaching, and in learning.

Should you feel our classroom interactions do not reflect an environment of civility and respect, you are encouraged to meet with your instructor during office hours to discuss your concern. Additional resources for expression of concern or requesting support include the Dean of Students office and staff (5-6757), the UI Counseling & Testing Center's confidential services (5-6716), or the UI Office of Human Rights, Access, & Inclusion (5-4285).

Times and places; papers and tests.

MMBB 488/ 588 Genetic Engineering

Genetics is an analysis of the blueprints that shape and form every virus, cell, and organism. This course will present you with an introduction to how this knowledge is being applied to reveal, and in some cases alter, these blueprints.

The course will follow the outline I am giving below; it will not follow the chapters in the book. In order to prepare yourself for each lecture you will need to become familiar with the handouts that I give you and with the notes of previous classes. I will specify chapters of Glick, Pasternak, and Patten (Molecular Biotechnology 4th edition) so that you may obtain a brief overview of the methods and goals of genetic engineering. Note that this text has a useful glossary. **USE IT.** This book was chosen to bring us to a common ground by mid-semester, after which, we will be discussing things that will only be found in class notes, and in research papers. A number of topics in genetics, biology, and biotechnology that are reviewed in the text will not be reviewed in the lectures because you have had them in earlier courses. I will only always welcome questions if any of these are not familiar to you, but the **decision to ask what they are or mean is yours.**

This is a provisional schedule for the course. Each section is intended to be covered over the course of one week, however some sections will likely be compressed in order to accommodate others, hopefully one suggested by members of the class.

TOOLS AND TERMS

WEEK OF 22 AUGUST

Section 1 (defining goals; model systems; basic function, properties, and organization of the gene; some characteristic differences between organisms; useful properties of bacteriophages, restriction enzymes, nucleic acid purification)

READ Glick *et al.*, pp. 1-46 and 3 sections of handout entitled, **Essential vocabulary, A quick review of the manipulation of nucleic acids, and Factors that distinguish translation, transcription, and replication.**

WEEKS OF 29 AUGUST-5 SEPTEMBER:

Section 2 (transformation of bacterial cells; selectable vs. screenable markers; plasmids; vectors and cosmids; genomic libraries)

READ Glick *et al.*, pp. 47-80, 222-228, 501-505. Read 3 sections of this handout entitled, **Basic features of the lac operon for genetic engineering, A sampling of selectable and screenable markers for transgenic organisms, and Rules and processes governing the appearance of a phenotype.**

WEEK OF 12 SEPTEMBER:

Section 3 (simple screening techniques; cDNA synthesis; gene isolation in bacteria).

READ Glick *et al.*, pp. 80-97, 164-172, and handout entitled, **Gene expression is determined by the following factors.**

WEEK OF 19 SEPTEMBER :

Section 4 (DNA transfer to eukaryotic cells; development of simple cloning vehicles for yeast, animals, and plants; fate of DNA that has been introduced into cells.)

READ Glick *et al.*, pp. 195-201, 240-253, 271-275, 278-279 and the handout entitled, **Representative methods for transforming different organisms.**

GENE ISOLATION AND EDITING

Week of 26 September:

Section 5 (Science and the analysis of scientific papers).

READ HANDOUTS AND ESPECIALLY BE SURE TO READ PAPERS INDICATED DURING CLASS

WEEKS OF 3-10 OCTOBER:

Section 6 (Sophisticated screening techniques: gene isolation through chromosome walking, transposons, subtractive libraries).

READ Glick *et al.*, pp. 205-208 and 5 sections of handout entitled, **With all the tools at our disposal, why do we still use classical genetics? Applying genetics to molecules, Classification of Mutations, Things to Keep in Mind When Recombination Seems Hopelessly Confusing, and Using a piece of DNA as a genetic marker.**

WEEKS OF 17-24 OCTOBER:

section 7 (Gene expression, requirements for transcription, designing promoters; conditional gene expression).

READ Glick *et al.*, pp. 723-738, 741-747.

Week of 31 October:

Section 8 (Exploiting the genetic code, value of introns in maximizing gene expression, optimizing conditions for translation).

READ INDICATED PAPERS and section of handout entitled, **A quick and dirty overview of translation and Use of different arginine codons in different organisms.**

WEEK OF 7 NOVEMBER

section 9 (protein targeting).

READ Indicated papers.

APPLIED GENETICS

WEEK OF 14 NOVEMBER

section 10 (The effects of gene addition)

READ Glick *et al.*, pp., 282-287, and 2 handouts in this packet entitled, **Operating Instructions for a Genetic Engineers and Establishing causality between transgene and phenotype.**

WEEK OF 22 NOVEMBER:

Fall Break

WEEK OF 28 NOVEMBER:

Sections 11-12 (The effects of gene removal; targeted recombination; sense and antisense; ribozymes).

READ HANDOUTS, Glick *et al.*, pp. 426-436, 747-748, 845-868, AND INDICATED PAPERS

WEEK OF 5 DECEMBER:

Section 13 (Understanding the unexpected: methylation, position effects, and transcriptional interference)

READ INDICATED PAPERS, Glick *et al.*, pp. 440-43, and handout entitled, **Characterizing transgenic organisms.**

(if time permits:

sections 12-13 (The effects of gene removal; targeted recombination; sense and antisense; ribozymes).

READ HANDOUTS AND INDICATED PAPERS

1. Where are we meeting?:

Class this year will be held in Life Science 163 Mondays and Wednesdays from 3:30-4:50.

You may find this time slot more challenging than the course material.

2. how am I going to be evaluated?

You will be given a number of quizzes and tests throughout the semester. **Every quiz and exam will include additional questions that must be answered by students in 588.**

Your final grade will be based on your performance in **4 areas**:

- **First**, the sum of the points earned from the 5-7 quizzes that will be given to you over the course of the semester,

The quizzes will be closed book. Each will be based either on the handouts and the text readings, or on 1-2 papers that will be given out previously. I will be testing you on either the reason certain controls or experiments were done, or on the conclusions of the paper. Questions will emphasize *the reasons things are done and the factors that determine success or failure* rather than specific details concerning the amounts of the materials used.

- **Second**, the points earned from a closed book exam to be given on **22 September**,

- **Third**, the points earned from a cumulative, open book final,

This will be “open book” meaning you can bring and use your notes, text, and any papers you feel would help you.

- **Fourth**, the points earned from a series of short questionnaires that will be handed out days when there is no quiz. You will **get** 1 pt for every one you turn in, and **lose** 3 points for every one you don't unless you tell me before class that you will not be coming.

Nota bene: Every quiz and exam will include additional questions that must be answered by students in **588**.

3. How do I figure out how I am doing in the class?

At the end of the course, all earned points will be summed and scaled. 488 and 588 students will be scaled independently. If you are interested in assessing how you are doing at any time, I will be providing a breakdown on how you did relative to others in the class with each test. The quizzes will be summed at the end of the course and weighted as a single test.

I have office hours following each class, but you are welcome to come by to talk with me about this or about the course any time. My office and lab are in rm 164b/ 161 Life Science Bldg. Early mornings are the best time to catch me.

A bit of advise

Recommendations for succeeding in this course

First and last, read this handout, and give the information some thought.

- *Considerate la vostra semenza:*

Fatti non foste a viver come bruti, Ma per segue virtute e conoscenza.

("You were not made to live as beasts but to follow virtue and knowledge.") — Dante

-The textbook will help you understand the underlying rules of molecular biology. The assigned readings will remind you of the salient features of different organisms and their genes, and at the same time, provide complementary presentations of some of the topics of molecular biology that I will be emphasizing in lecture. The greater portion of the course, though, will deal with how these ideas are being applied to solve problems and so will depend on specific research publications. It is strongly advised that you learn to read and analyze scientific papers: this will be one of the skills I hope you will acquire in this course. A bibliography of all experiments discussed in lecture is provided for this purpose. However, everything you are expected to know will come from the summaries of this body of work presented in lecture and in the handouts. Hopefully, my lectures and the discussion you initiate based on them will provide all of the information you need to understand the state of the art today.

-Take detailed notes, don't depend on your memory or on tape-recordings. The activity of writing helps imprint the information on your memory better than any passive behavior.

-Review these notes no later than one week after the lecture. You will find the techniques and organisms mentioned in one session serve as the foundation for later ones. *The best review is to form a study-group with one or more other students and take turns summarizing each conclusion and the experiments that demonstrated it.* Speaking out-loud helps fix the information in your memory; listening to yourself trying to reason helps reveal areas where you are uncertain or confused; your audience serves as an editor to point out errors. Fill in the sentences, conclusions, or unanswered questions that may have occurred to you in your notes. Get the answer to those questions as soon as possible by going to the original article, to follow-up articles, or by raising the question in the next class.

Always remember during these discussions to ask each other for the evidence supporting each conclusion. Reading a paper or listening to a scientific lecture is like playing any challenging game. Examine every fact, whether it comes from Nature or from your co-workers, as you would examine every move or play in a game. Close scrutiny may reveal weaknesses arising from incorrect assumptions, faulty execution, or incorrect or overly extrapolated interpretations. Every experiment has limits, some acknowledged, some not. Good scientists are not rigid. In order to plan the next experiment, in order to link previously distinct bodies of data, they attempt to extrapolate beyond the limits of the interpretations, so long as this does not violate a physical or biological "law". They may even go beyond these laws if experiments put the laws in doubt. It is the job of the reader to search for and find these flaws. But it is not enough to deny the observations or

interpretations just because they go beyond your expectations or imagination: it is fine to be skeptical but you must come up with solid reasons stating why the experiment is flawed and what you require to accept it.

-learn how to integrate information from all model systems; you may never work with E. coli or tobacco, nevertheless, what is valid in one organism is more often than not, partly valid in others. And solutions that solve one problem are worth using to solve problems occurring in others.

-Although it may seem obvious to you, I have found that a significant number of people beginning this class have trouble extracting generalities from examples and conversing, appreciating the need to offer examples and exceptions for each generality. You are expected to see both the specific and the universal in each case that we cover so if you have trouble, or are uncertain whether your interpretation is correct, be sure and talk to me.

-I will generally be available discussions after class (the best time to ask questions), or at 5:00 any day of the week in rm 230 and in a rm across from it in Ag Science Building. Once the locations have been settled, I will send everyone more definitive information. Don't let this uncertainty prevent you from seeking further information from me. **You will generally be welcome at other times as well. Come by or arrange an appointment.**

A review of words and concepts

Essential vocabulary

What is a gene?

The copy of nucleic acid that encodes information for another molecule in the cell. Genes consist of a transcribed region and the flanking sequences needed for the appropriate transcription of that region (a **promoter**, **terminator**, and regulatory control elements). The majority of transcripts are then translated, but some genes only make an RNA that then becomes a component (such as *ribosomal RNA* or *transfer RNA*) for an enzyme encoded by another gene, while others make an RNA to regulate other RNAs in some way. Note that all genes are inherited (passed on from a cell to its daughters) and all cellular genes are encoded in DNA. A given DNA molecule (a plasmid or chromosome) may have tens of thousands of genes on it, or only a few.

What is genetics?

The scientific study of 1) the transmission of information from generation to generation and 2) the processes by which the information is used. Geneticists isolate mutations in particular processes, then map their locations, and finally characterize their actions until the blueprints directing the appearance of a particular trait have been deciphered. There is a subtle but very important distinction between genetics and **molecular biology**. Molecular biology grew out of a blend of genetics, physical chemistry, and biochemistry. It focuses on the isolation of genes and their products, sometimes on their sequence, and very much on how specific nucleotide or amino acid sequences produce each biologically important character of the molecule. Molecular biology and genetics overlap to a large extent, but one can isolate mutants and use them to explain the way specific genes contribute to a phenotype without ever isolating the actual gene product. Conversely, one can study how a molecule works without ever isolating a mutant. **Molecular genetics** encompasses and blurs the distinction between both disciplines.

What is a promoter?

A promoter is the cis-acting sequence at the 5' end of a gene that binds **RNA polymerase** and positions it so it can transcribe (synthesize an RNA copy of) that gene. Promoters are "directional", that is, like a sentence of a language, they are only meaningful if read in one orientation (5' to 3'). They direct RNA polymerase to transcribe sequences along one strand beginning at the 3' end of the promoter. Promoter sequences will not be present in the messenger. (For those who like analogies: the RNA polymerase "jumps" from the promoter to the transcribed region of the gene like a diver jumping from the diving board to the pool). A "bi-directional" promoter is in fact two adjacent or overlapping promoters that are directing transcription from opposite strands of the DNA. **Note:** in this course, you must distinguish between the promoter and all cis-acting sequences that regulate the activity of the promoter: **transcriptional regulatory sequences** such as **enhancers**, **silencers**, **operators**, and **positive-regulators** will not be considered parts of the promoter.

What is meant by transcription?

Transcription is the enzymatic reaction carried out by RNA polymerase using free **rNTPs** and a double-stranded DNA template. In order to be transcribed, that template must contain a specific nucleotide sequence where RNA polymerase can bind, a sequence within 10-13 (for prokaryotes) or 30-35 bases (for eukaryotes) of the 3' end to that promoter where RNA polymerase can begin copying the template, and a sequence farther downstream directing polymerase to terminate the transcript. Like virtually everything we will discuss, the processes of transcription, translation, and replication are the same in all organisms but the sequences involved and the proteins that recognize them can differ radically from genera to genera and kingdom to kingdom.

What is meant by translation?

Translation is the enzymatic process carried out by **ribosomes** using a) **aminoacyl-tRNAs**, b) several different auxiliary proteins, c) a **messenger RNA**, and d) GTP. The **genetic code** specifies 20 common amino acids and 1 very unusual one. Each amino acid incorporated into proteins has a corresponding **aminoacyl-tRNA synthetase** and family of tRNAs. (note: Enzymatic charging of the tRNA with its appropriate amino acid requires ATP).

What is meant by replication?

Replication is the enzymatic reaction carried out by DNA polymerase using a) free **dNTPs**, b) a number of auxiliary proteins, and c) a single or double-stranded DNA template. In bacteria, most DNA molecules begin replication at specific sites present 1-2 times per chromosome. These specific points are called **origins of replication**. DNA synthesis requires a **primer**-generally an RNA or DNA chain ending in OH- and based paired to a part of the template. The enzyme begins adding new bases to the free hydroxyl group of the primer so that they will be complementary to the sequence of the unpaired portion of the template. Thus, if the original DNA molecule had a gap in it, replication could fill that gap in so that every base was paired again with its complement.

What do we mean when we use the word transformation?

The most common use of the term refers to the introduction (or the method of introduction) of a new gene into a cell or organism. One of the sources of confusion associated with this word is that it applies whether or not the gene is retained for any period of time. In general, constant selection, or integration of the foreign DNA in a resident replicon ensures **stable** transformation. However, some experiments investigate the effect of a gene shortly after it is introduced (**transient expression**), before organisms have had time to integrate or lose the introduced gene.

Note that the term transfection only applies to the introduction of a virus, or a viral vector, into a cell by artificial means. Transfection is not synonymous with transformation.

What is a vector?

This is another term that can confuse beginners. In its strictest sense, it is a tool to introduce a gene into a cell or organism. However, for many people it has become synonymous with a plasmid or autonomously replicating episome. This ignores the fact that some vectors, especially ones in common use in plant and animal transformation, are transposon- or virus-like sequences.

Many vectors are just that—a nucleic acid structure that contains, in addition to all of the genetic elements essential for its own replication and maintenance, a restriction site where we can clone the gene our attention is focused on. Unless the gene being cloned has all of the expression signals necessary for function in the recipient host, it will not work there. cDNA sequences, by the very nature of their origin, will not contain any of the sequences required for transcription. In order to express cDNAs, which are essentially little more than the protein-coding region of a gene, one uses an expression vector that has all of the features of a vector plus one or more cloning sites designed expressly to allow the DNA sequences inserted there to be transcribed and translated in a host cell. In these sorts of constructs, the cloning site has a promoter at its 5' end, and a transcription terminator downstream. Some of these expression sites have been built with additional elements such as short open-reading frames so that the cloned gene would be translated with a new N- or C-terminus. These modifications to the cloned gene ensure it is either translated efficiently or targeted to a defined cellular or extra-cellular compartment where it would be most useful.

Finally, remember that vectors are limited by the capacities of the genes from which they are built. A vector for *E. coli* is built from *E. coli* promoters and origins of replication. If, therefore, you want it to work anywhere else, you will have to add additional sequences that are functional in the alternative host. For example, if we wish to build an expression vector which would allow us to test random cDNAs for their ability to complement yeast mutants, we might start with a plasmid having 1) a ColE1 origin and 2) tetracycline-resistance gene so that we can initially generate our library in *E. coli*. Because this is an easy host to transform, it is used to recover the greatest possible number of independent clones. However, because the assays will be done ultimately in yeast, we would have to modify this bacterial vector with the addition of 3) a yeast origin (for example from the 2 μ circle), 4) a selectable marker which could work in yeast, and 5) an expression site built from yeast terminators and promoters to ensure that full-length cDNAs would be expressed in the final host.

The fundamentals of genetics

Factors that distinguish translation, transcription, and replication

	replication (also called "DNA synthesis")	transcription (also called "RNA synthesis")	translation (also called "protein synthesis")
Where is it done in prokaryotes?	In the cytoplasm	In the cytoplasm	In the cytoplasm
Where is it done in eukaryotes?	In the nucleus, mitochondria, and chloroplasts	In the nucleus, mitochondria, and chloroplasts	In the cytoplasm , mitochondria, and chloroplasts
What is the general name of the enzyme involved?	DNA polymerase (a breakdown of this word means: <i>enzyme that makes DNA macromolecules</i>)	RNA polymerase (a breakdown of this word means: <i>enzyme that makes RNA macromolecules</i>)	a ribosome
What does this enzyme use as a source of information to synthesize the product?	A pre-existing double-stranded DNA molecule	A pre-existing double-stranded DNA molecule	An mRNA molecule
What are the precursors of the reaction?	dNTPs	rNTPs	aminoacyl-tRNAs, GTP
What are the products of the reaction?	molecules of DNA (plasmids and chromosomes) as large as the original template	RNA of varying sizes depending on the size of the gene	proteins (also called polypeptides) of varying size according to size of ORF
What cis-acting sequences determine where the enzyme begins and ends its designated reaction?	Origins of replication (also called replication origins); terminators of replication (also called termini)*	Promoters and transcriptional terminators*	Ribosome binding sites (also called Shine-Dalgarno sequences in prokaryotes and Kozak sequences in eukaryotes), start codons, stop codons (also called termination codons*)

*Note: to avoid confusion, we only use the term "terminator" when referring to transcription. In order to avoid confusion, we must use the **complete phrase**: "terminator of replication" when referring to DNA synthesis, and "termination codon" or "stop codon" when referring to the making of protein.

For practical purposes, we cannot go into all of the details that distinguish each of the biological kingdoms. With apologies to other bacteria, cyanobacteria, streptomycetes, and the two kingdoms of archaea, we will consider E. coli to represent all prokaryotes. For the moment we will also lump all eukaryotes (fungi, protists, animals, and plants). The relevant exceptions will become apparent as the course progresses.

Noteworthy features that distinguish **gene expression** in prokaryotes and eukaryotes

	bacteria	eukaryotes
number of types of RNA polymerases in common use	1	3 (although only one transcribes mRNA)
frequency of introns	very rare	very common
translation in same compartment as transcription?	yes (ribosomes and chromosomes are in the same cell compartment)	no (80S ribosomes are in the cytoplasm while transcription is in the nucleoplasm. However, transcription and translation is coupled in organelles that have evolved from prokaryotic endosymbionts)
importance of poly-A tail	generally absent, very short when present, role unknown	generally present, very important for mRNA stability
importance of 5' cap on mRNA	not used by bacteria	generally essential for initiation of translation
internal initiation of translation on mRNA?	yes (and messengers often polycistronic)	very infrequent (messengers <u>almost</u> exclusively monocistronic)
use of initiator codons other than AUG	In E. coli, <10% of genes: >10% in gram-positive bacteria	<u>almost</u> exclusively AUG

Rules and processes governing the appearance of a phenotype.

1. All phenotypic traits, of both prokaryotes and eukaryotes, are ultimately under the control of genes that are encoded in DNA, or in the case of some virus', RNA. DNA is expressed when it is copied into RNA which in some cases is subsequently translated into protein.

nota bene: Although all phenotypes are under genetic control, do not assume phenotypic traits are only controlled by genes: the phenotype *now* can be a reflection of past growth conditions or inherited developmental information^①

(a phenomenon called epigenetic control).

Also, do not assume all genes must be translated in order to be expressed. MicroRNAs and related regulators are surprisingly abundant in multicellular eukaryotes.

Information by definition is transferable order. A gene contains the information for the formation of an RNA template; this RNA contains the information for the order of polymerization of amino acids into a protein. A protein can contain the information for the assembly of a macromolecular structure. A cell can contain the developmental information for what the adjacent cell will become within an organism. An organism can tell you you're not leaving the house without combing your hair. Ad infinitum. Each transfer of information allows entropy to decrease locally.

2. Although almost every cell of a multicellular organism has the same collection of genes, one must not assume every cell expresses the same sets of genes at the same time. The selective silencing and expression of different genes is the key to the process of development that allows a cell to produce a daughter which is genetically identical yet phenotypically different.

3. Cellular DNA is organized into functionally independent chromosomes and sometimes episomes. Each requires one or more origins of replication. Virus' can replicate extrachromosomally or by integrating into host replicons.

4. Genes can be extracted from cells and put back.

These techniques are duplicating or exploiting natural gene transfer processes found in all kingdoms such as virus infection and transfection, conjugation, nuclear transfer, cell fusion, and genetic colonization.

① Genes don't always work well in heterologous organisms, and so interspecific gene transfer is very rare. The fun of working in the laboratory is that there are ways to improve this artificially.

5. DNA can be cut into pieces at selected sites by restriction enzymes and put back together by DNA ligases.

6. In order to be expressed, a gene needs i) a promoter where regulatory molecules and DNA-dependent RNA polymerase binds, ii) a sequence where RNA polymerization begins, iii) a sequence where ribosomes can bind, iv) a start codon where protein synthesis begins then proceeds through an open reading frame (ORF) and stops at a designated termination codon, and finally, v) a stop-of-transcription sequence.

7. Prokaryotic gene expression differs from eukaryotic expression in that most eukaryotic mRNAs have long poly-A tails while prokaryotic ones have none, or short ones; eukaryotic messengers generally have caps; eukaryotic ribosomes can only efficiently translate the first ORF of a series (except with the aid of gene-specific supplementary factors). In addition, many eukaryotic mRNAs are "spliced" (a specific form of processing) from long RNAs containing fragments of discontinuous coding sequence. By contrast, few prokaryotic genes are spliced.

8. Proteins are made according to a (nearly) universal genetic code. Nevertheless, different organisms, or even different genes within an organism may use different synonymous codons preferentially.

9. (Not all regulation is transcriptional. Nevertheless,) It is possible to alter the expression of many genes by altering transcriptional control elements. Phenotypic traits can be altered by purposefully increasing or decreasing the expression of a gene (or the accumulation of its gene product) in the appropriate cell type, and/or at the appropriate time.

10. Most organisms share a common set of metabolic pathways although the enzymes of different organisms may differ in activity, stability, or post-translational regulation because of differences in the protein sequence that have evolved to suit each organism's unique niche. It is possible to change the physiological capabilities of an organism by giving it a biochemically distinct isozyme taken from a heterologous species, or by replacing its native enzyme with the evolutionarily divergent isozyme.

11. Eukaryotes are divided into compartments-nuclei, mitochondria, chloroplasts, vacuoles, golgi, ER, etc. The first three structures have their own genomes. These encode and express genes that may remain in their compartment or go to another. The location of a protein may be altered by the addition or deletion of specific amino acid "targeting" sequences, or by altering the compartment where the corresponding gene is located.

12. The perversity of the universe tends to increase. Anticipate and look for the unexpected. When it occurs, learn from the result how to make more effective gene modifications and how to use other genes to correct the deficiencies.

Gene expression is determined by the following factors:

process	determinant	location	does the efficiency of this process differ from gene to gene in eukaryotes?
transcription initiation	promoter	nucleus	greatly ¹
capping	(capping done automatically by RNA polymerase)	nucleus	no
poly-A addition and termination	polyA addition site transcriptional terminator	nucleus	somewhat ²
splicing	intron sequences	nucleus	somewhat
nuclear transport of RNA	possibly non-specific	nucleus	no (?)
mRNA turnover	sequences related to (AUUUA) _n	nucleus and cytoplasm	somewhat
translation initiation	nucleotide context of the start codon	cytoplasm	somewhat
efficiency of decoding	codon usage, concentration of individual aa-tRNAs	cytoplasm	somewhat
protein folding	polypeptide sequence and concentration of individual chaperones	cytoplasm	somewhat
protein turnover	specific and nonspecific proteases	cytoplasm	greatly

¹ several orders of magnitude

² less than an order of magnitude

Note carefully: the efficiency of gene expression is determined separately from regulated gene expression. For our purposes, the determinants of efficiency act all the time, whereas regulation implies that expression is sometimes low, and later high, or visa versa.

A QUICK AND DIRTY OVERVIEW OF TRANSLATION

(What follows represents the minimum information you should know about translation. If you are unfamiliar with any term, please see me.)

For each of the 20 most common amino acids, there is at least one aminoacyl-tRNA synthetase which discriminates between its appropriate amino acid and tRNA(s) and all of the other inappropriate molecules. The enzyme then forms a high energy bond between the two types of molecules to generate an aminoacyl-tRNA. Each tRNA has an anticodon, which, because of the wobble rules and the restrictions imposed by modifications made to bases adjacent to the anticodon of each tRNA, is capable of binding to between one and three of the codons assigned to each amino acid. There are cases where the same codon may be recognized by different isoaccepting tRNAs. However, protein synthesis is heavily biased against the translation of codons by aa-tRNAs of different families: there is very little ambiguity in the genetic code. All of the steps involving the decoding of the message are conducted on the ribosome with the aid of its proteins and RNAs, additional initiation and elongation factors, and the driving energy of GTP hydrolysis.

Some amino acids (for example, tryptophan and methionine) may use one codon, others as many as six (e.g. leucine, serine, and arginine). In addition to the 61 codons coding for amino acids, there are three generally assigned to termination factors. When these triplets (stop codons) appear in the A-site, termination factors (which are proteins, not nucleic acids) direct the peptidyl-transferase of the ribosome to transfer the growing peptide to water instead of another amino acid, and the newly synthesized protein is released. In very special circumstances which we will not discuss in this course, one of these stop codons can code for a 21st amino acid, selenocysteine.

The code is read in adjacent, non-overlapping triplets within a context determined by the first codon of the open reading frame. Most ORFs begin with AUG, but some use GUG, CUG, or even AUU. Regardless of the way each is read during subsequent elongation steps (methionine, valine, leucine, and isoleucine, respectively), here at the very beginning, each of these four codons is translated as methionine by an initiator tRNA. Prokaryotic protein synthesis actually begins with N-formylmethionine, but the formyl group is added only after the methionine has been joined to the tRNA. Later, after translation has begun, the formyl or the formylmethionine group may be removed.

The translation of an orf ends with a stop codon. The proteins that bind there during the process of translation direct the ribosome to transfer the peptide to water instead of to a new aminoacyl-tRNA. After protein synthesis has been completed, additional modifications may be made to change the genetically coded amino acids to other forms. This can include conversion of proline to hydroxyproline, threonine to phosphothreonine, lysine to ϵ -methyllysine, etc. If you aren't familiar with the way these changes in structure might alter the chemistry of the group, review amino acid structure and function using a basic biochemistry text.

Use of different arginine codons in different organisms

arginine codon	E. coli	Bacillus	humans	mouse	budding yeast	fission yeast	tobacco	maize
CGA	5.2	11.6	10.6	11.3	6.7	5.2	11.1	6.1
CGG	37.2	18.4	<u>21.2</u>	19.8	6.7	4.6	8.8	30.5
CGU	7.7	10	20.3	17.4	2.2	2.9	5.1	15.2
CGC	40.6	23	8.7	9.1	7.3	15.9	17.9	11.1
AGA	3.5	28.6	18.6	<u>21.2</u>	89.3	51.3	29	9.6
AGG	2.4	8.2	20.5	<u>21.2</u>	1.8	17.9	28.1	27.5

The most frequently used codon in each species is underlined; the least frequently one is italicized.

Nota bene: Each family of codons is "decoded" by a family of aminoacyl-tRNAs. There is only one tryptophan codon and many organisms have only one tRNA that can be charged with tryptophan and used to decode that codon. On the other hand, the 6 arginine codons require at least 3 (and not infrequently more) different tRNA species to decode them. Each species is encoded by its own gene. Each gene is transcribed at a determined rate to meet the demands for that species. Since different organisms "prefer" different arginine codons, each organism transcribes each arginine tRNA gene at a different level.

Each time a tRNA is used to decode a codon, it transfers its amino acid to the growing peptide and then has to be recharged before it can be reused. If the demand for a species exceeds the number of charged tRNAs available, translation of every messenger needing that particular species halts until some becomes available. Weakly expressed genes rarely tax the cell's capabilities to supply the charged tRNAs. On the other hand, strongly expressed genes containing rare combinations of codons can. In these cases, the amount of protein made from each of these messengers containing rare codons is much less if translation has to repeatedly stop and wait for more tRNAs to be charged.

With all the tools at our disposal, why do we still use classical genetics?

Every trait of an organism has a genetic component. One identifies the genes responsible for these characters by isolating mutations in which that trait is altered. Mutations produce phenotypes by blocking or short-circuiting steps of underlying biochemical, physiological, or developmental pathway governing each facet of the trait—when it appears, where it appears, and how it appears when given the appropriate environment or internal stimulus. Rather than becoming obsolete with the introduction of newer biochemical and “molecular” methodologies, genetic analysis has become more important. It is now possible to use our knowledge of genetics to isolate or characterize an unknown gene product. For example, suppose you have a newly isolated bacterial, fungal, or mouse gene but do not know its function in the cell. In this event, you can generate a specific mutation in the corresponding locus by recombining artificially-modified portions of the cloned DNA with the chromosomal copies to produce an inactive allele and in some cases, a novel phenotype. These “knock-out” experiments will become more and more important as the various Genome-sequencing projects begin to generate more and more unclassifiable cDNA sequences.

Mapping genes and performing complementation analysis tells us whether a gene or an identified genetic activity (a phenotype) is different from all others known to date. For example, a phenotype may depend on several proteins interacting to form a single enzyme. A mutation in the gene for one of these subunits may be biochemically indistinguishable from a mutation in any of the other ones. Mapping studies, however, would show each mutation was located in a different part of the genome. Mapping a mutation within a gene (“fine structure mapping”) is also one tool to determine which functional component of a gene is defective or altered (something sequence data can’t always provide. Why?).

Complementation studies would show whether the gene product is freely diffusible (hence trans-acting) or restricted to action in one dimension (cis-acting). The same data could also be used to verify that one mutation was or was not in the same place in the genome as another.

A third tool of genetics is the isolation of suppressors. Extragenic suppressors of a mutated gene can be used to identify cellular products that interact or cooperate with the suppressed gene to produce a normal organism, or identify genes with similar capacities to the mutated one.

Identifying a ts, cs, or information-suppressible allele of a locus indicates whether the suppressed gene encodes a protein. Any conditional mutation can be used to determine the critical period when a gene product is needed to produce a normal organism, not just when it was made.

Classification of Mutations

This is not meant to be either diagnostic or encyclopedic. It is meant to stimulate you to think about what you learned as undergraduates in genetics and to provide examples of the types of mutations that will be mentioned in this course. As you look over this table, keep in mind one thing: for the most part, molecular and biochemical techniques can show you what molecules are responsible for a phenotype, but not which is most important or actually limiting the process. Mutational analysis, by contrast, can tell you which gene is most important, but not necessarily what its function is. The best work of the last 30 years has succeeded because these two approaches complement each other so well.

type	site	common molecular characteristics	magnitude of phenotype	more probable suppressors
base change ¹	promoter or regulatory element	change amount of mRNA synthesized; specific activity of protein that is made not changed	change amount of mRNA synthesized generally or only in specific growth environments	reversion; base change very nearby creating substitute promoter or regulatory element
base change	open reading frame	no change in transcription; no protein or truncated protein (if nonsense mutation or frame shift); normal amount of protein with altered activity (if missense mutation). Missense mutations can also cause proteins to be conditionally active.	mild to severe (can be conditional on temperature, pH, presence of exotic chemicals like DMSO or ethanol)	reversion; nonsense suppressor; missense suppressor; frame-shift suppressor; compensatory change in protein elsewhere to restore activity to poorly functioning protein. Can get compensatory change in another protein if it interacts with this one.
deletion	promoter and regulatory element	change amount of mRNA synthesized (more severely than produced by base change); specific activity of protein that is made not changed	severe	rare
deletion	open reading frame	often no visible transcripts, or ones with altered size or half-life; the little product that is made often inactive and/or unstable	severe	very rare (generally only possible where another gene can be mutated to assume the function of the lost gene).
insertion of DNA	promoter and regulatory element	change amount of mRNA synthe-	change amount of mRNA	excision/ deletion of DNA

		sized (some sequences can introduce new promoters giving unregulated gene expression); specific activity of protein that is made not changed	synthesized	
insertion of DNA	open reading frame	often no visible transcripts, nor active protein	severe	excision/ deletion of DNA

1 base changes can be neutral if they a) do not alter sequences essential for transcription, translation, replication, etc., or b) alter them by substituting one base for a functionally equivalent one. For example, changing GTG to GTT may have no effect if the only function of this sequence is to code for valine within the translated portion of the gene. Changing the codon to CTG, which would substitute leucine for valine, might be nearly neutral depending on the importance of the amino acid in the completed protein.

Things to Keep in Mind When Recombination Seems Hopelessly Confusing.

First, remember recombination is a perfectly natural activity. Every organism can do it, and most do it some time in their lives. In bacteria, it occurs whenever cells mate, or when plasmids come into a cell with sequences identical to those on the host chromosome, or when transposons move, or when integrative phages, well...*integrate*. It sometimes even happens between repetitive sequences (such as IS elements) on the chromosome. Although it may be more sophisticated than the ligation reactions in our Eppendorf tubes, recombination is still nothing more than the cutting and joining of DNA molecules within the cell.

There are two types of recombination to keep in mind in this course. One is termed homologous recombination. It occurs between identical or nearly identical sequences. Sequences as small as 10-30 bases are enough for this reaction. In some cases, the enzymes doing the cutting-and-joining have no sequence specificity. They catalyze "general recombination". In other cases, the enzymes only work at precisely defined sequences. These catalyze "site-specific" recombination.

The other type of recombination is termed illegitimate, irregular, or, perhaps incorrectly, non-homologous recombination. In this process, superficially dissimilar sequences are joined. In fact, there are often several bases of similarity near the join sites, reflecting the recognition biases of the enzymes involved, where one sequence has recombined in his manner with another. These enzymes include some of those involved in general recombination, but not all, and may be helped by additional proteins that are normally not used for general recombination.

Every DNA molecule can undergo either process. The frequency with which each occurs depends on the activity of the recombinases and the probability that these enzymes will optimize base-pairing before cutting-and-pasting. In small genomes, illegitimate recombination events between freely diffusing DNA are less probable than homologous recombination events. As the genome size increases, the former is more and more favored unless the cell is particularly recombinationally active.

Its easy to provide enough homology for legitimate recombination to occur frequently, but as you will learn through reading or class notes, several tricks are needed to ensure your selection or screening operation will find these. These tricks serve one function: to reduce the chances that "transgenics" (resulting from

the incorporation of exogenously provided DNA in a particular genome) and "recombinants" (resulting from crossover events between different portions of a genome) made through nonhomologous recombination survive selection.

Each of the following terms defines a different type of recombination. In general, the enzymes involved are the same, and the products are predictable; only the substrates differ.

Intramolecular recombination

Single crossovers within a molecule generate circles. Thus, recombination between homologous repeats within a circular molecule leaves us with two circular molecules. Recombination between sequences within a linear molecule produces one circular molecule and one linear one.

Intermolecular recombination

When two circles, or one circle and one linear molecule, recombine **through a single crossover**, one gets One Big Circle, or One Big Linear molecule, respectively. On the other hand, when two linear molecules recombine, you get two linear molecules back. If one lacks an origin of replication, it is lost. Thus, if you have introduced a linear molecule into a cell, and recombination occurs between it and the chromosome, then a portion of the chromosome becomes part of a nonreplicating plasmid. You can guess what will happen next.

Multiple crossovers

Multiple recombination events are suppressed within the vicinity of the first one, probably by simple steric hindrance between the recombination complexes (it takes so many protein species to do this reaction right that you can easily see sites undergoing recombination by means of an electron microscope). However, new recombination events frequently happen upstream or downstream of the first event, perhaps occurring when the complex drifts from the first site after completing the reaction there. Even-number recombinations result in the simple exchange of sequences between the two points.

It is also possible for a sequence to be copied from one molecule into another by a process termed "gene conversion". Sometimes this new sequence is merely added to the target molecule, sometimes it replaces a sequence already there. The product looks like a recombination event but it isn't reciprocal: one sequence has been duplicated, and often, one sequence has been wholly erased.

Establishing causality between transgene and phenotype

Correlation between clones obtained and phenotype under investigation	Analysis needed to verify cloned gene is in fact responsible for the phenotype under investigation
very high; in rare cases, overexpression of a dissimilar gene suppresses phenotype	If complementation due to a genomic clone (big enough to encode more than 1 gene), than the responsible gene must be identified by testing subclones.
very high	The wild-type gene must be isolated and demonstrated to complement the mutation.
initially very low	Ultimately, fragments of N_{ω} must be tested by complementation, or shown always to coincide with sites where mutations cause the same phenotype.
low	Technique establishes correlation of expression with phenotype, but the magnitude of this contribution not known. Moreover, additional work may be required to identify which gene on a large genomic fragment corresponds to the observed hybridization signal. Not all relevant genes of a phenotype show differences in mRNA levels. This technique biased to identification of very well expressed transcripts.
unambiguous	When applied correctly, technique ensures you will find the desired gene if it is in the library. When not applied correctly, you will get the desired gene and several others which have similar nucleotide sequences. Additional techniques then required to discriminate correct from incorrect candidates.

Applying genetics to molecules

Comparison between several different approaches to gene identification and isolation.

Technique	Applicable to ¹ :	Description and requirements
Complementation	bacteria, fungi, (animals)	An appropriate ² genomic or expression library is used to complement a recessive phenotype. Transformation efficiency must be high, mutants must be first identified; selection or screening procedures must be unambiguous; reversion frequencies for mutation must be lower than the probability of transferring 1-2 copies of the screened genome.
Transposon, insertion, or T-DNA tagging	all organisms	Mutants are isolated using transposable elements or inserts of foreign DNA. The frequency with which DNA is inserted must be much higher than other forms of mutation in the organism. The inserted sequences must be cloned already so that they can be used as probes to pick up genomic clones. These inserts should be rare in the genome, preferably unique, so that the chances of picking up and identifying the copy in the sequence that is responsible for the phenotype is high.
Chromosome walking	all organisms	A phenotype is mapped as precisely as possible to a single locus. Cloned DNA fragments (N_0) from the same linkage group are used to find adjacent clones (N_1) lying between them and the targeted locus (N_0). Fragments of N_1 are used to find N_2 , and so on until N_n is reached.
Differential screening for correlated genes	all organisms	Genes expressed in specific tissues, at specific times, or under specific developmental or physiological conditions (state 1) are isolated by preparing RNA or cDNA probes to material from state 1 and state "other-than-1" and comparing which library clones each reveals. Modifications of this can be used to search protein populations instead of nucleic acid ones.
Screening by identity	all organisms	Gene probes from other organisms, or synthetic nucleotide probes designed to hybridize to sequences coding for specific peptides are used to screen cDNA or genomic libraries.

¹Organisms in parentheses denote those where the method is currently being used successfully, but requires much greater efforts.

²The genes transferred must be able to be expressed in the new host.

Using a piece of DNA as a genetic marker

Populations are divided geographically and/or behaviorally into partially isolated gene pools. As time passes, mutations (single-base changes, deletions, or insertions) arise within individuals in these pools. Each mutation creates an allele that potentially alters one or more of these characters. Some are passed to the next generation, others are lost through selection or stochastic processes (sampling effects).

The phenotypic differences between two individuals often represent differences in the sequence of the genes at independent loci. We can trace the transmission of each locus through into the next generation by looking for the associated trait. Since each locus is part of a chromosome, traits on the same chromosome tend to be transmitted together very frequently. Since most eukaryotes are diploid, one runs into cases where the particular allele on one chromosome differs from the allele on the other. Organisms with two different alleles at a single locus are said to be heterozygous. Most organisms are heterozygous at a high percentage (even a majority) of their loci.

Homologous chromosomes can undergo recombination leading to re-assortment of the combinations of distinguishing characters encoded on each homologue. The more frequently two particular loci are found together, the less likely that a recombination event will separate them. The more individuals you can test for each trait, the more precisely this distance can be known. A genetic map is merely a pictorial summary of the frequency with which each locus studied assort with its neighbors among the progeny of individuals heterozygous for those traits.

A useful genetic map should have uniformly spaced markers spanning each arm of each chromosome. Unfortunately, some genes are hard to map accurately because their phenotypes are not always visible in all offspring. Some phenotypes, for example, are only apparent at certain times during development, or not apparent unless the organism is grown in unusual conditions. Moreover, some of the allele-specific differences used to characterize a locus are recessive. As a result, phenotypically identical individuals may be genetically different and one is often forced to discard individuals that cannot be genotyped unambiguously from mapping experiments. In the case of other loci, no genetic variants have been detected yet, for example because mutations that alter the activity of the product produce a lethal phenotype.

In an effort to obtain an alternative set of genetic markers to incorporate into an unfolding genetic map, geneticists have turned to the use of "restriction site polymorphisms" to map, and ultimately isolate genes. These sorts of markers arise when a given mutation in the DNA erases or adds a restriction site, or changes the size of a restriction fragment through the loss or addition of sequences. Although such changes might not be visible macroscopically, they can be seen if they have occurred in the vicinity of a molecularly characterized portion of the genome. For example, imagine you have a piece of DNA encoding a gene, a part of a gene, or an uncharacterized piece of a chromosome selected at random from a library of a plant genome. This fragment can be used as a probe to

investigate through Southern analysis whether individuals of different subpopulations have differences in this particular region of their genome. In order to do this, you would prepare genomic DNA from several individuals, cut each with several different restriction enzymes, prepare genomic blots, and then probe them.

Imagine, for example, you cut the DNA with EcoRI, and after hybridizing with your probe, found that your probe hybridized to a 2kb fragment in all individuals. This would mean that in the chromosomal neighborhood of the sequence homologous to your probe, the sequence of these 2 EcoRI sites had not been altered, nor had any DNA integrated or been deleted between them.

Next, suppose you looked at the same DNA cut with BamHI and probed with the same probe. Suppose in this lane, the DNA of one plant had a band of 2 kb, while the others had ones of 10 kb. This difference could mean something had been lost in the first individual to bring these BamHI sites closer, or something (without Bam sites) had been added to this portion of the genome of the other individual to push the restriction sites farther apart. Alternatively, something with a BamHI site had been added to individual 1, or a base change had created a new site in that plant that allowed the 10 kb fragment to be cut in two (with probe homology on only one band). Finally, it is possible that a base change had erased a site from that region of the other individuals.

Any change in the local restriction map for a fragment is referred to as a restriction fragment length polymorphism (RFLP). RFLPs map and segregate like any genetic trait. They also have certain advantages: 1) they need not be in a gene, or may be due to a single base change that is otherwise phenotypically invisible. Such changes can be selectively neutral and as a result, can accumulate between populations without altering the competitive abilities of the population. 2) RFLPs can be analyzed using DNA from any part of the organism, not merely one tissue such as flowers that may appear only once in a plant's life. 3) Their detection isn't confounded by environmental effects or masked by the expression of other genes. 4) In a similar way, they are detected equally well in homozygotes or heterozygotes of that locus, and in fact provide a means to distinguish one from the other. Finally, 5) that piece of DNA you used as a probe against uncloned DNA can be used to clone the entire region from a library. The fragment you used as a probe may in and of itself be useless, but by providing a probe for genetic analysis, it could be used to obtain an important one.

RFLP maps for fruitflies, nematodes, mice, humans, rice, potato, maize, pine trees, etc., etc., etc., are now well developed or in progress. Many genes for medically or agronomically important traits have been cloned.

Operating Instructions for a Genetic Engineer

1. Choose a problem.

The **genotype** represents a set of instructions or blueprints for the formation and functioning of nearly all parts of an organism. Unfortunately, we don't know yet how to "read" all of these instructions. Most of the genetic engineering projects being done today are designed to teach the meaning of this genetic text. We proceed by making specific changes in the blueprints for an organism in order to reveal what part of a phenotype is changed by the malfunctioning of a particular gene. Once the genetic determinants for a particular phenotype have been understood, one can begin to modify organisms to give them new or augmented traits with medical, industrial, or commercial value.

It is encouraging that enough has been learned so that biodegradable plastics can be made in plants, human genetic diseases can be studied in modified mice, and the size and habitat range of commercially important fish can be increased.

2. Investigate the problem: find out what is known about the molecules contributing to the phenotype.

The genetic engineer must first and last be an eclectic biologist, not restricted to one discipline such as bacteriology, mammalogy, or plant breeding. Every process in every cell involves interactions between molecules, and similar molecules are often used in very different organisms. In order to give yourself the greatest chance to succeed in a project, you must find out all that is known about the phenotype you are investigating. This most definitely means you must search the literature for what is known about that trait in each of the major phylogenetic groups-archaea, bacteria, eukaryotes (fungi, protozoa, animals, plants, alga). In some cases, people have investigated the problems in very different but complementary ways in each of the different groups. You may save yourself years of effort by using a solution found by someone else in an obscure organism.

3. Choose an organism.

If you are reasonably certain a modification will work, you could try working in the organism that attracted your attention to the problem. On the other hand, this organism may be difficult to transform or slow to grow. Thus, if you want to try something new, you won't want to spend a few years trying to transform a difficult system only to find the planned change didn't work or had unforeseen consequences. In general, you will be best off testing your idea out in an appropriate model system first.

(model systems are organisms chosen because they have rapid life cycles, are well understood genetically, and easy to modify because someone has already developed the needed genetic tools.

These include E. coli, budding yeast, fruit flies, zebra fish, mice, Arabidopsis, and tobacco)

4. Choose a gene.

Traits are determined by the expression of one or more enzymes or structural genes. If you wish to modify that trait, you must have a copy of the gene. Your first thought might be that you need the gene from the organism you wish to engineer. But as the previous step emphasized, don't restrict yourself through ignorance of alternatives. Before you set out to clone a particular version of a gene, ask yourself whether a satisfactory homologue has already been isolated from some other beast. You may wish to use that copy rather than spend time finding the one in your organism. If the enzyme activity is equivalent, it may do regardless of its source. On the other hand, if the work depends on obtaining a new gene, ask yourself "which version of the gene would be best?" Since different organisms have slightly different genes, and in some cases, several possible alleles of that gene, you can choose which protein would likely show the greatest stability or turn-over number to suit your specifications. After all, evolution has provided us with millions of versions of each gene, and some are truly unique in their biochemical properties because they evolved in unusual circumstances. If you must isolate a gene, why not go for Nature's best, rather than the most commonplace.

5. Ensure it will work.

There are subtle differences in the molecular processes in different phyla and kingdoms that can work against the naive engineer. Different groups of organisms use slightly different sequences for transcription, translation, and turn-over of each gene and gene product. If you have ever had a course on evolution or comparative physiology, you will find that enormously useful in genetic engineering. The farther apart two organisms are evolutionarily, the more you need to modify the genes of one to ensure they will work when put in the other. To use a trivial analogy, both IBM and Mac computers can process documents in English, but that doesn't mean words typed using IBM software can be read by the software used on the Apple computer. In the same way, the promoters/ RNA polymerases and ribosome binding sites/ ribosomes of prokaryotes differ from eukaryotic or archeal equivalents far too much in a few key details to allow genes from one group to be transcribed and translated in another. However, just as you can reformat a paper to read in a different computer operating system, you can make a few base changes in a gene so that it will no longer work in its original host but will in a different one.

Keeping all this in mind, you now see that when you begin to plan a project you must 1) select a suitable promoter and transcription terminator, 2) select a suitable start of translation and a suitable stop of translation for your gene, 3) remove any gene processing elements like introns that might not work in the new host and add other gene processing details like a sequence specifying polyadenylation that might be necessary in the new host, and 4) verify that the codon usage in the new gene is similar enough to the codon usage in the new host.

6. Find a way to introduce the gene stably in the new host

In order to isolate a gene from its original host, it must be cloned onto a vector capable of replicating in or integrating into an existing replicon of either *E. coli* or yeast. However, this cloned copy is an intermediate form for most of the experiments you are likely to do. To be useful, it must be then transferred to a vector that either replicates in or integrates into the genome of the host you intend to change. Any transfer of DNA from test-tube to intermediate host, or from intermediate to final host is going to be a rare event ($<10^{-3}$ successes/ cell). In some cases, it will be a very rare event ($<10^{-7}$ successes/ cell). And it will be lower still if you lack an efficient way to introduce the DNA into the new host. You need a simple way to distinguish successfully transformed from untransformed cells, for example, a selectable marker that lets transformed cells live when all others die.

These markers are often chimeric genes on their own, built from open-reading frames from one organism fused to expression signals from another. Each chimeric gene has the same kinds of host-range limitations as the gene you have modified and wish to introduce. Each part of the selectable marker gene may fail in the new host so before you do anything, you must verify whether you are able to introduce this there, and determine whether it works well enough to ensure it can be used to select for transformation. Determining whether this marker can function in a novel environment is generally harder and more time-consuming than anything you will do later. However, once you have gotten it to work, you can introduce this selectable marker together with your gene of interest and obtain an organism making both.

7. Verify you have added a gene and altered nothing inadvertently.

There are number of transformation methods, each with its own pitfalls. In general, you can't be sure you have introduced a new gene until you prove it is there by several independent tests. You'll get a separate hand-out detailing these tests. The guiding principle is don't assume you have succeeded because you have a resistant organism, or something that appears positive by PCR. In addition, don't assume the genes are working correctly just because the DNA is there. Gene rearrangements, mutations, or inactivation are common in some systems or when some transformation techniques are used (see point 8).

8. Identify the best transformed individuals in the population that you have produced, but also determine the average performance of your transgenic population.

In general, genes introduced on plasmids or integrated by homologous recombination behave most predictably. These options for maintaining the new genes, unfortunately, aren't always available. The alternative methods-non-homologous recombination, or use of integrative vectors derived from transposons, can give different results depending on where the DNA linked to the genome. Thus, each transformant can

express the new genes to different levels, if at all. You must analyze a number (5-15 often) of independently generated transformants in order to determine the average/most typical effect and the best effect for your purposes.

Absorb the next section. It matters.

Characterizing transgenic organisms

"The perversity of the universe tends to increase"

-First Law of Quantitative Theology

"...in direct proportion to your proximity to fulfilling your life's ambition."

-Finagle's Unwelcome Commentary

There are an awful lot of ways to fool yourself into believing the most favorable interpretation of an experimental result. Since eukaryotic transformation processes can proceed so slowly, it is easy to understand why one wants to believe that it has been successful. Unfortunately, errors are made; failures are possible. There is an exceptionally long history of enthusiastic reports of successful transformation techniques that have never been successfully repeated. These were not the result of fraud, but rather naiveté and self-deception. To reduce the chance that you or a friend join these ranks, do more than one test before assuming an organism has been transformed successfully. The following are a set of critical tests you must make before allowing yourself to become convinced.

I) Whenever possible, use vectors with more than one selectable marker, or with both screenable and selectable markers. In this way, one marker can be used to verify what the other one produced and show that the selected organisms acquired their new traits from your vector and not by mutation. For example, if your vector contained two markers, and you selected with marker A, then every true transformant should later be shown to have marker B, even though it was not used initially.

II) When vectors with multiple markers are not available, determine the efficiency with which you can get the phenotype of the selected marker when the vector alone is introduced into the new host, and the frequency with which you could get similar phenotypes without introducing any new genes (this is the spontaneous mutation frequency for that trait). If the transfer or stabilization efficiency is low compared to the mutation frequency, you might not be able to distinguish mutations from transformation.

III) Verify transformation by physical detection of DNA. This may mean re-isolating the plasmid if you are introducing genes on a small replicon into bacteria or yeast, or detecting gene integration via Southern analysis.

IV) When doing Southern analysis, always include positive and negative controls that allow you to demonstrate you could have detected the new gene if it was there. In general, this means, include one lane of digested DNA from untransformed cells (the negative control), and one with DNA from untransformed cells that has been spiked with the DNA you tried to introduce (the positive control). In this latter case: **be absolutely sure the amount of the gene-of-interest you add when spiking the sample is approximately equal to the amount of the gene expected in the average transformed line.** This generally means the equivalent of one-two copies /haploid genome. (If the gene you are adding is 1000 basepairs long, and the genome you are adding it to is 10^7 basepairs, then that gene would be 10^{-4} of the transgenic genome. Therefore, if you load 10 μ g DNA from each

transformed candidate to each lane, the control lane should be spiked with 1 ng of the pure gene-of-interest.)

V) Determine the approximate copy number and the organization of the introduced gene in the transformed genome by using at least three different enzymes. The first enzyme should cut out an internal segment of the gene you wish to see expressed. The more intense the hybridization signal to this band, the more copies you have introduced. The second enzyme should cut once and only once in the middle of the gene or vector. In the simplest cases, the number of bands detected after hybridization will then approximate two times the number of copies integrated. The third digest should use an enzyme that does not cut in the vector. The number of bands detected in this lane equals the number of independent integration sites. The results from analysis of these three digests should give a consistent estimate of copy number, gene organization-after-transformation, and whether all copies integrated separately or in tandem.

VI) Whenever possible determine the number of co-segregating integration loci transmissible meiotically. The number of independent loci is not equal to the number of copies integrated. An organism might have five copies integrated in tandem in a single locus, or five scattered over five chromosomes. The consequences for expression, genetic stability, and uniformity of inheritance are very different in the two cases.

Tables and tools

A quick review of the manipulation of nucleic acids

DNA and RNA are both quite hydrophilic, as are carbohydrates, and many small molecules like nucleotides, amino acids, small peptides, organic acids, etc. By contrast, lipids and most proteins are hydrophobic. Most protocols for the isolation of DNA depend initially on this dichotomy.

Cells are broken, generally as gently as possible, in the presence of detergents, ion chelating agents, and a buffer. One tries to be gentle when preparing DNA because large molecules are easily sheared at interfaces between liquids, or between liquids and the solid walls of containers and pipettes. The detergents prevent molecules from aggregating and trapping nucleic acids. Chelating agents are included to sequester divalent ions (especially Mg^{+2} and Ca^{+2}) which are needed for the activity of many enzymes. By removing them, you inhibit (at least partially) some of the destructive enzymes that are released when you open the cells and which can degrade the DNA and RNA you are preparing. There are a number of additional precautions (which can be found in more detailed protocols) that can be taken to inhibit these enzymes. The ionic detergents noted above can also help inactivate proteins by unraveling some of their hydrophobic regions. The buffer helps maintain an appropriate pH since extremely basic solutions can denature DNA and hydrolyze RNA.

Buffered phenol is then added to the mix where it forms a second, and denser, phase. The hydrophobic portions of proteins tend to go into this phase, leaving their hydrophilic portions sticking up into the interface. Unfortunately, polyadenylic acid also tends to go into the phenol phase so if you were to discard this phase at this step in order to get rid of the proteins, you'd lose some of your mRNA. To avoid this, chloroform is added to increase the density of the hydrophobic layer and reduce its water content. Now, with the proper precautions, you can remove the lower phase and interface and so eliminate most of the cellular proteins and lipids. There are then additional fractionation techniques to separate RNA from DNA. You will hear about some later, or can read of them in the materials and methods sections of many of the papers, as well as in most biochemistry textbooks. Carbohydrates generally follow your every move, but luckily have few deleterious effects on nucleic acids or the enzymes used in nucleic acid studies. Newly available resins that bind nucleic acids preferentially according to their hydrophobicity and charge can be used to eliminate residual impurities. Others can be lost from mRNA preparations when mRNA is separated from rRNA, tRNA, and single-stranded DNA.

The isolated material can be stored in your freezer indefinitely. DNA might be used for making libraries or for restriction analysis. For example, DNA can be cleaved into discrete fragments by treatment with restriction enzymes. One then separates these according to size by drawing them through agarose gels by means of an electric field. The fragments of genomic DNA form uniform "smears" beginning with the largest ones at the origin and the smallest at the opposite end of the gel. This is then transferred to a binding matrix such as nylon or nitrocellulose where fragments are fixed by one of a variety of methods and hybridized with radioactive or otherwise labeled fragments of particular genes. Hybridization is visualized using photographic film and the presence or absence of an appropriately migrating band constitutes failure or success of the **Southern analysis**. If RNA is used (northern analysis), no digestion is done so the size of the final band denotes the size of the messenger RNA.

Highly purified messenger is usually prepared from eukaryotic cells by binding mRNA to columns or paper that is coated with oligo-dT. The oligo rA tails of

the transcripts form stable duplexes with the oligo-dT. Degraded mRNA, or any of the aforementioned molecules, washes through. You can then elute the mRNA from these solid supports. This is standard practice for the first stages of cDNA synthesis.

By the by. Don't let any of these terms or concepts slip by you. They are central (but generally not mentioned) aspects of most of the experiments discussed in this course.

A sampling of selectable and screenable markers for transgenic organisms

Enzyme (substrate)	Original Host	Use	Most common Applications
neomycin phosphotransferase (kan,neo, G418)	bacteria	selection, qea ¹	bacteria, fungi, animals, plants
hygromycin phosphotransferase (hyg)	bacteria	selection,qea	fungi, animals, plants
adenosine deaminase (9-β-D-xylofuranosyl adenine)	mammals	selection	primarily in some mutant mammalian cell lines
dihydrofolate reductase (mtx)	mammals	selection	primarily in mammals, some use in plants
thymidine kinase (5-bromo deoxythymidine)	mammals	selection,qea	animals
xanthine-guanine phosphoribosyl transferase (aminopterin +mycophenolic acid)	bacteria	selection,qea	animals
chloramphenicol acetyltransferase (cam)	bacteria	qea	bacteria, fungi, plants, animals (good selectable marker in bacteria)
luciferase (luciferin)	insect (firefly)	qea	bacteria, plants, animals
β-galactosidase (ONPG, X-gal)	bacteria	qea, histochemical	bacterial mutants, fungi, animals
β-glucuronidase (p-nitrophenyl gluc-uronide, X-gluc, 4-methyl-umbelliferyl-β-D-glucuronide)	bacteria	qea, histochemical	fungi, plants
green fluorescent protein (mutants isolated that fluoresce blue or red)	ctenophores; coelenterates	qea, histochemical	bacteria, fungi, plants and animals
aminoglycoside phosphotransferase (sm)	bacteria	selectable marker (in bacteria), histochemical (in plants)	(see adjacent column)

¹qea: quantifiable enzyme activity; activities useful for screening.

Note: the choice of system depends on whether the organism under investigation has a similar endogenous activity. β-galactosidase, for example, can't be used in lac⁺ bacteria, or in plants. Similarly, some organisms are relatively resistant to some selective agents making it difficult to distinguish transgenic organisms from untransformed ones.

A CLONERS VIEW OF CRITICAL FEATURES OF LAMBDA

This is not a complete description of the virus and the genes that regulate it. That information will be presented in lecture, or may be found on pages 80-83 of the text. The purpose of the following paragraphs is to present the barest minimum needed to use lambda as a cloning or expression vehicle.

LAMBDA infects *Escherichia coli*. In this host, the phage propagates either lytically or as a lysogen. When wild-type virus infects wild-type hosts, the phage generally replicates lytically. This can be demonstrated by mixing a few phage with a vast excess of bacteria and spreading the culture uniformly on nutrient agar. After 20 hr of incubation at 37°, circular plaques form in an otherwise uniform bacterial lawn. Each plaque is a nearly bacteria-free zone resulting from the repeated cycles of infection, replication, release, and new infection beginning with the first infected cell. Depending on the specific phage and host, each λ plaque may contain up to 10^5 phages. These plaques are turbid due to the presence of a few thousand very small colonies growing in them. Each microcolony is the result of an infection where the incoming virus lysogenized (integrated into and became part of the host chromosome) instead of multiplying. Only two or three genes of the virus are expressed within the lysogen, but one of them, C_I , makes the host immune to superinfection by LAMBDA so that the lysogen can form a colony in spite of the sea of viruses around it.

One can grow a lysogen like any other *E. coli* stock in liquid medium, and for all practical purposes, it would appear to be a normal bacterial line. However, if the culture is exposed to UV light, or to chemicals that damage DNA, the hitherto quiescent virus excises and begins replicating until it lyses the host.

Mutations can alter this life-cycle. If there is no repressor (C_I), or C_{II} or C_{III} protein, or if there are mutations in the operators where C_I binds, the plaques appear clear because the phages cannot lysogenize. Phage with mutant operators are called vir mutants. They can even infect lysogens because immunity depends on C_I -mediated repression. Other mutations, in phage or host, only allow the virus to lysogenize so no plaques form at all.

LAMBDA encodes the genes for a DNA packaging system capable of enclosing in a viral coat its own chromosome (appr. 49 kb), or other DNA of the correct size and organization. Given the right DNA substrate and active proteins prepared separately from infected cells, this packaging can be done in vitro. The DNA substrate must

have two copies of the cos sites 45-51 kb apart. The intervening DNA may be the virus itself, or DNA from any organism that you may wish to insert instead.

LAMBDA contains several, highly regulated, very active promoters. P_L for example can be cloned separately and used to express genes in *E. coli*. Its great usefulness stems from its mode of regulation by the C_1 gene. There is a commonly used allele of the repressor called C_{1857} which is almost as active as wild-type at 28-30°, and almost inactive at 40-43°. This gene has been cloned from λ and inserted into non-lysogens. Any gene expressed in these strains under the control of P_L or P_R can thus be turned on and off at will by shifting the temperature on a growing culture.

λ life cycle:

a) phage adsorbs to lamB.

-note: to ensure full induction of lamB, bacteria are grown to exponential phase in maltose-supplemented medium.

b) DNA is injected and circularizes at cos sites.

c) N and cro are made. N allows C_{II} to be made, C_{II} allows repressor to be made.

-either

d') C_I turns off synthesis of N and cro and turns on its own synthesis. C_I shuts off transcription from P_L and P_R . Without sufficient late gene expression, phage lysogenizes in primary attP site of the bacterial chromosome. Phages remain quiescent until induced by damage to DNA. Damaged DNA induces a host protease which cleaves specific host proteins and also λ repressor. This then allows phage to excise, and enter the beginning of the lytic cycle (d" below).

-or

d") proteases destroy C_{II} so that no more repressor is made, while N and cro still are. Late genes (such as genes P and Q which are involved in rolling circle replication of DNA) are produced rather than being shut off, and phage replicates as an episome. DNA is packaged, lysis gene is produced (the S gene product), the host pops, releasing viruses.

-a detailed review of these processes can be found in many biochemistry texts, almost all genetics textbooks, and all basic texts on molecular biology. Good descriptions are also found in *A Genetic Switch*, M. Ptashne. Call number QR 342 P82.

Properties of some representative selective agents

antibiotic (abbrev)	mode of action	common mode of resistance (gene designation)	most common use
ampicillin, carbenicillin (amp, cbn)	inhibits formation of peptidoglycan cross links; kills only growing bacterial cells	β -lactamase cleaves molecule (<u>bla</u> , or less correctly, <u>amp^r</u> , <u>cbn^r</u>)	bacteria
chloramphenicol (cam or CM)	inhibits peptidyltransferase of bacterial and mitochondrial ribosomes	chloramphenicol acetyltransferase adds acetyl group to inactivate molecule (<u>cat</u>)	bacteria
fusidic acid	inhibits EF-G to block translocation	mutations in <u>fusA</u> or <u>fusB</u>	bacteria
kanamycin (kan)	inhibits binding of aa-tRNA to decoding site of ribosome and peptidyl-tRNA translocation; primarily effective against 70s-type ribosomes *	neomycin phosphotransferase (<u>npt-II</u>) phosphorylates anti-biotic to inactivate it; also active against neomycin and G418	bacteria . eukaryotes
G418	acts like kanamycin, but primarily against 80s ribosomes	neomycin phosphotransferase (<u>npt-II</u>)	eukaryotes
hygromycin (hyg)	acts like G418	hygromycin phosphotransferase phosphorylates antibiotic to inactivate (<u>hpt</u>)	eukaryotes
methatrexate (mtx)	inhibits dihydrofolate reductase	mtx-resistant dihydrofolate reductase (<u>dfr</u>)	mammals
rifampicin (rif)	bacteriostatic; inhibits binding of β -subunit of bacterial RNA polymerase to holoenzyme so RNA synthesis blocked	mutated subunit does not bind antibiotic (<u>rif^r</u> , <u>rpoB</u>)	bacteria

spectinomycin (spc)	bacteriostatic; inhibits peptidyl-tRNA translocation on 70s ribosomes	aminoglycoside phosphotransferase (<u>apt</u> , <u>str/spc^r</u>) or aminoglycoside adenylyltransferase (<u>aad</u>) (active against strepto-mycin also) modifies antibiotic to inactivate it; mutation in protein S5 of ribosome which prevents binding (<u>rpsE</u> , <u>spc^r</u>)	bacteria
streptomycin (str or SM)	inhibits translation of genetic code by 70S ribosomes	aminoglycoside phospho-transferase (<u>apt</u>), aminoglycoside adenylyltransferase (<u>aad</u>), or mutation in S12 of ribosome to prevent binding (<u>rpsL</u> , <u>str^r</u>)	bacteria
tetracycline (tet)	inhibits binding of aa-tRNA to bacterial ribosomes	<u>tet^r</u> gene decreases permeability of cell	bacteria

*You might remember that the sedimentation rate of prokaryotic ribosomes has been described as 70S. Organellar ribosomes evolved from such structures. Although some organellar ribosomes sediment somewhat faster than some bacterial ones, both are structurally similar and sensitive to the same antibiotics. Eukaryotes have an additional set of cytoplasmic ribosomes termed 80S that translate nuclear-encode messengers. their antibiotic sensitivities differ as noted here.

A Field Guide to the Common Transposons of the Laboratory Bench

name	approx. size (bp)	structure ^a	phenotypic markers	original host	trans-genic range
Tn1	5000	IR-tnp-rep-bla-IR	amp ^R	enteric bacteria	enteric bacteria
Tn5	5700	IR-npt-str-ble-IR	kan ^R (also str/spc ^R and ble ^R in some hosts)	enteric bacteria	many gram negative bacteria ^c
Tn10	9300	IR-tet-IR	tet ^R	enteric bacteria	enteric bacteria
Ty1	6300	DR-gag-pol-DR	none	yeast	yeast
P	2700	IR-tnp-IR	none	dipterans	dipterans ^b
Ac	4500	IR-tnp-IR	none	maize	all plants tested ^c

^a abbreviations: IR, inverted repeat; DR, direct repeat; tnp, transposase; reg, repressor of transposition; gag, polyprotein which is cleaved into various structural proteins for encapsulation; pol, reverse transcriptase, integrase; ble, bleomycin resistance; npt, neomycin phosphotransferase; str, streptomycin resistance; bla, β -lactamase; tet, tetracycline resistance.

^b unless modifications of splice sites and promoter have been made.

^c regulation and frequency not always normal in transgenic hosts.

T-DNA mediated transformation of plants

By far the most common vehicle used to transform dicotyledonous plants such as petunia, tobacco, tomato, potato, etc exploits the machinery of the Ti-plasmid of Agrobacterium tumefaciens. This plasmid contains a number of genes that carry out the operations all plasmids must do: they replicate the plasmid DNA, they help it conjugate from bacterium to bacterium, and they help the bacterium by conferring resistance so some phages and providing the means to catabolize various sugars and amino acids found in the soil in the vicinity of plants. However, there are a number of additional features on the 180-210 kb episome that are unlike anything else known in nature. Principal among them is a region called the T-DNA that encodes between 9-13 genes (depending on the particular plasmid considered) and is bounded by 25 bp direct repeats. Nearby, distributed over 40 kb of the plasmid, lie a large number of genes forming the vir (virulence)-region of the plasmid.

All of the T-DNA genes, and most of the genes of the vir-region are transcriptionally silent throughout the life of the bacterium. However, genes are expressed when the bacterium is exposed to specific sugars and phenolic compounds that are released from plants that have been wounded either as they pushed their way through hard soil, or as they were brushed against by a passing hiker. One of these vir genes gene makes a nick in one strand of each of the T-DNA repeats. Next, other genes begin replicating the T-DNA from these nicks, mimicking processes normally carried out by analogous gene products during the process of plasmid transfer from bacterium to bacterium. The single-stranded copy that is made has a vir-encoded protein (virD2) bound covalently at its 5' end. It is thought the naked DNA then leaves the cell with the help of other vir proteins and is then coated extracellularly by several hundred copies of another vir protein (virE2) so that the DNA is protected, like a coat protein protects a virus, when the DNA moves from bacterium to plant. In the plant, the DNA diffuses to the nucleus, becomes double-stranded, and integrates at random sites in the chromosomes. The enzymes at work during this integrative process have not been identified but it is thought the pilot protein attached to the end of the T-DNA, and several host DNA-repair enzymes, are involved. Note that the T-DNA is immobile after it has integrated: all the genes needed for transfer were left behind in the bacterium, as were portions of the flanking repeats where the vir-proteins acted.

The T-DNA genes become active in their eukaryotic host. Three of these genes encode hormone-synthesizing enzymes, and at least two make hormone modifying proteins. The summed effect of these is the

formation of either a plant tumor or "crown gall", or a mass of ectopic roots. One or more of the other gene products takes normal amino acids and organic acids or sugars to synthesize compounds found nowhere else in nature (referred to as opines) and unmetabolizable by the plant. These are then exported to the tumor surface where daughters of the infecting bacterium sit in wait. The Ti-plasmid contains genes specifically geared to metabolize these new compounds so these cells can grow. The result is that the bacterium has "colonized" the plant and created a source of food for itself that no other organism can use. It has made its own niche.

In the last decade, genetic engineers have adapted this capacity for DNA transfer from bacterium to plant for their own ends. The tumor-promoting genes have been removed and been replaced by whatever gene the scientist desires to insert in the plant. Generally, a selectable marker gene is included. Now when the T-DNA transfers, it carries with it a new set of traits. In most cases, the transformed cells can then be cultured and regenerated into a plant.

While some pathogens have rather restricted host-ranges, A. tumefaciens is able to infect most broad-leafed plants, and some gymnosperms and lilies. This has made it as important as lambda to agricultural genetic engineering. It can transfer DNA to monocots, but successful integration of genes in the host has not been demonstrated unambiguously.

Representative methods for transforming different organisms.

This is a compilation of some of the methods in common use, or under investigation to transfer DNA to various organisms. The list is not exhaustive. Numbers represent commonly reported frequencies (shown, where appropriate, in clones recovered/ μg DNA used). Be careful when comparing these methods because some required DNA to integrate through illegitimate (also referred to as nonhomologous or irregular) recombination (#), while others used much more efficient specialized vectors (•) capable of recombination or replication. ?=data not available.

	E. coli	fungi	animals	plants
phages	1×10^7 - 2×10^9 (•)	? ^a	?	?
plasmids	(10^{-5} - 5×10^{-1} /recipient) (•)	(3×10^{-8} - 7×10^{-2} /recipient) (•) ^b	?	?
Agrobacterium	(using <u>tra</u> functions, $1 - 2 \times 10^{-3}$ /recipient cell) (•)	$3.3 (10^{-3})$?	(generally 0.1-10% of protoplasts) (•)
Ca ²⁺ ppt.	$10^6 - 10^8$ (•)	?	$10^3 - 10^4$ (#) $10^4 - 10^5$ (•)	?
PEG	?	10-100 (#) $1 - 2 \times 10^4$ (•)	?	10-100 (#)
electroporation	$10^4 - 10^5$ (#) 10^9 (•)	$10^6 - 10^7$ (•)	$5 \times 10^3 - 3 \times 10^5$ (•)	10-100 (#)
particle gun	?	?	(0.02-0.05% of cells) (#) (•)	(up to 0.1% of cells) (#)
microinjection	?	?	(5-25% of oocytes injected) (#)	?
lipofection	?	?	$1 - 7 \times 10^5$ (•)	?

My apologies to those interested in protozoa. Microinjection as well as techniques for transforming fungi are generally applicable to them.

?^a: Both of these vectors depend on recognition of receptors in the native hosts. Introducing these receptors into new genetic backgrounds might make these organisms into hosts. ^b: In this rather remarkable example (See reference list for this section), E. coli plasmids have been shown to conjugate to yeast. There are no known conjugative plasmids in eukaryotes.

What follows is a highly condensed version of how one of the best characterized and most utilized operons in *E. coli* works. Understanding this system in its natural state will help you understand what ingenuity was required to adapt it to use in other bacteria and especially in eukaryotes. I presume you have heard much of this information before, but if you have never read the actual studies which lead to our present understanding, you've missed delving into one of the most enlightening bodies of scientific literature in modern biology. It employed sophisticated genetics, jerry-rigged biochemistry, and some awfully clever model building & testing by some of the founders of molecular biology including F. Jacob, J. Monod, and W. Gilbert. I've included a list of some of the relevant papers below; I wish we could go over these and more.

Lets first review one method by which mutants can be isolated. One can mutagenize a bacterial stock and plate them on medium with glucose as a carbon source. Then, when the colonies have begun to grow, one can lay a piece of velvet on the plate so that a few bacteria from each colony are picked up by it. If you then press this cloth onto a new plate containing lactose, some of the bacteria will be transferred and thus will form new colonies in exactly the same position as on the original plate. A few of the original colonies, however, will be mutants that are unable to utilize the lactose. Without any other carbon source, these bacteria will not grow on the replica plates. You can then go back to the master and pick those colonies to save. Once you've got some of these mutants, you can clone each allele and sequence it. Sometimes the gene will prove to be inactivated because a base in the protein-coding region has been changed so the protein is inactive. Other mutations, however, will be in those bases of the regulatory regions of the gene critical for expression. For example, one can pick up mutations in the -10 and -35 regions of the promoter, yet not find mutations in the bases inbetween. Clearly, every position is susceptible to mutation so why should the mutations obtained be so clustered? On reflection, you will see that mutations in any base that did not affect the lactose-utilization phenotype would not have been picked up in this screen because the hosts would have been phenotypically normal.

E. coli has four genes involved in the catabolism of lactose. One is β -galactosidase that you will hear frequently about in class. It is an enzyme of 1021 aa encoded by the lacZ gene. This gene is the first of a three member operon. Immediately following

it is the lactose permease (lacY), which brings the sugar into the cell, and the galactose transacetylase (lacA), whose importance in the cell is somewhat uncertain. Upstream from this operon is an independently expressed lac repressor, encoded

by lacI. The lacI protein is made constitutively. It binds to the adjacent region between it's own gene and the lac operon. This region consists of a promoter and operator functioning somewhat like that of $\lambda_{P_{L}O_L}^*$ or $\lambda_{P_{R}O_R}$. This repressor:DNA complex ensures the operon is almost never transcribed.

The repressor binds random DNA sequences about 10^8 times weaker than operator sequences. Nevertheless, the protein is always in equilibrium between these two classes of sites and so, based on its estimated concentrations in the cell and the size of the genome, it is concluded that less than 2% of the repressor is bound to the operator. When DNA replicates, or cell volume or ionic milieu fluctuate in response to normal environmental changes, the repressor may drift to one of these nonoperator sequences allowing a little bit of operon expression. This permits a few molecules of the permease to be made and become incorporated in the cell membrane. When lactose becomes available, some of it can enter the cell through these rare permease molecules. In the cell, the trace amounts of β -galactosidase convert these to allolactose. Allolactose is the true inducer of the operon. Allolactose binds the repressor, regardless of whether it is bound to operator or nonoperator, and shifts the equilibrium by reducing the affinity of the protein for the operator 1000-fold. It is also possible to de-repress the operon with analogs of allolactose, especially isopropylthiogalactoside (IPTG) which does not require the permease in order to enter the cell. Without repressor to block it, RNA polymerase can now begin transcription.

During the growth of the bacterium, transcription can be enhanced by a number of pleiotropic effectors: molecules that serve as intracellular signals to coordinate otherwise independently expressed genes. The two major factors that act like this are guanosine 5"-diphosphate-3"-diphosphate (ppGpp or "magic spot") and 3',5'-cyclic AMP (cAMP). The former enhances the frequency of transcription of this operon more than 10 fold. The latter binds to a cAMP-binding protein (=catabolite repressor protein, CRP) and this complex in turn binds a specific site of the regulatory region of the lac operon where it enhances expression as much as 50 times, probably by an effect on the activity of RNA polymerase at the time of transcription initiation. cAMP levels are low when cells are growing at their maximum rate as when there is glucose in the medium. As a consequence, cells grown with lactose and glucose aren't fully induced for β -galactosidase (this effect is called "catabolite repression").

Magic spot is made when cells enter N,PO₄, or S starvation conditions. Normally, the key element of this response is to

turn off as many pathways as possible in order to conserve matériel (the stringent response). rRNA and tRNA synthesis is the first to shut down, and apparently this is in part done by shifting the affinity of RNA polymerase from a preference for stable RNA promoters to one for mRNA. If there is no lactose around, the additional polymerase molecules sitting in wait on the promoter have no effect. If lactose is around, and only nitrogen, mostly needed for amino acid synthesis, is lacking, then the extra mRNA produced might compensate for the reduced rates of protein synthesis per molecule. The full relevance of this particular part of the process is still a puzzle.

The lac promoter and some mutant derivatives of it are among the most commonly used expression systems in *E. coli*. They have been cloned into both phage- and plasmid-based vectors. One variant promoter called ptac (a hybrid promoter made up of the ideal -10 region of plac and the -35 of the promoter of the tryptophan operon) can raise the level of the foreign gene product to as much as 20-30% of the total soluble protein in the cell. This level can also be achieved using P_L.

When the operon is cloned on high-copy plasmids, it is sometimes advisable to use hosts that overproduce the lac repressor since plasmids can be so abundant that the number of **operators** in the cell becomes greater than the number of effective repressor molecules. As a result, some of the copies of the lac promoter are always on.

Note one other important attribute of the lac system. There are, in addition to artificial inducers such as IPTG, artificial substrates. One is ONPG (o-nitrophenol β -galactoside) which is colorless until cleaved, and yellow afterwards. By measuring the amount of ONPG hydrolyzed/min/mg protein, you can obtain an estimate of the amount of enzyme. By determining the amount of enzyme, you can estimate the strength of a promoter. Thus, fusions of β -galactosidase to new promoters has allowed people to show changes in promoter activity after induction, or over the course of development. One other substrate you will encounter throughout this course is 5-bromo-4-chloro-3-indolyl- β -galactoside (X-gal). This is colorless. When cleaved, indole is released which precipitates in insoluble blue crystals. Thus, any cell making β -galactosidase and stained with this substrate turns blue while non-expressing cells remain clear. This provides a histochemical assay to determine the specificity of different promoter: β -galactosidase fusions that have been introduced into animals or fungi.

Amann, E., et al., 1983. Vectors bearing a hybrid trp-lac promoter useful for regulated expression of cloned genes in Escherichia coli. *Gene* 25: 167-178.

Gibert, W. and Müller-Hill, B., 1966. Isolation of the lac-repressor. *PNAS* 56: 1891-1898.

Jacob, F. and Monod, J., 1961. Genetic regulatory mechanisms in the synthesis of proteins. *J.Mol.Biol.* 3: 318-356.

Lin, S.-Y. and Riggs, A., 1975. The general affinity of lac repressor for E. coli DNA: implications for gene regulation in prokaryotes and eukaryotes. *Cell* 4: 107-111.

Pardee, A., B., et al., 1959. The genetic control and cytoplasmic expression of "inducibility" in the synthesis of β -galactosidase by E. coli. *J. Mol. Biol.* 1: 165-178.

Primakoff, P. and Artz, S.W., 1979. Positive control of lac operon expression in vitro by guanosine 5'-diphosphate 3'-diphosphate. *PNAS* 76: 1726-1730.

*Note: one of the less fortunate ambiguities in common molecular symbology is the use of "p" to denote plasmid (e.g. pUC18) and promoter (e.g. plac). Look for the letters that follow the p. Plasmids are not underlined, gene designations are, and the promoter is usually followed by the name of the gene from which it originated.