

ЦЕЛИ, ПРИНЦИПЫ И ПОНЯТИЯ МОЛЕКУЛЯРНОЙ ЭВОЛЮЦИИ

1.1. Задачи молекулярной эволюции как науки

Молекулярная эволюция как наука включает в себя два направления исследований. Ее целями являются изучение законов изменения наследственной информации в живых системах, включая доклеточные и клеточные формы жизни, и изучение истории развития жизни на Земле, установление родственных отношений между формами жизни, их генетическим материалом — филогении форм жизни (phylogeny). Первое направление включает в себя изучение частоты и других характеристик эволюционных изменений в макромолекулах — нуклеиновых кислотах и белках, а также механизмов и причин, определяющих эти изменения. Вторая область исследований молекулярной эволюции, направленная на реконструкцию эволюционного процесса развития форм жизни на Земле, установление родственных связей между формами жизни, создание их эволюционной классификации, называется молекулярной филогенетикой (phylogenetics). Традиционно изучение пребиотической эволюции, происхождения жизни, также понимается как задача молекулярной эволюции. Однако механизмы пребиотической эволюции принципиально отличаются от законов эволюции живых систем, требуют специального рассмотрения и, как правило, описываются отдельно.

1.2. Нуклеотидные последовательности

У всех живых систем, включая доклеточные и клеточные формы жизни, наследственная, или генетическая, информация содержится в геноме, представленном молекулами нуклеиновых кислот. У подавляющего большинства форм жизни генетическая информация передается от поколения к поколению в виде молекул дезоксирибонуклеиновых кислот (ДНК). Ряд вирусов является исключением из этого правила и передает генетическую информацию в виде рибонуклеиновых кислот (РНК) — РНК-содержащие вирусы.

Молекула нуклеиновой кислоты представляет собой линейную полимерную цепочку, состоящую из мономеров — нуклеотидов. В эволюционном анализе различия в химическом строении

между дезоксинуклеотидами и нуклеотидами часто для простоты описания игнорируют, говорят о нуклеотидах вообще. Каждый нуклеотид состоит из азотистого основания, сахара (дезоксирибозы — у ДНК, рибозы — у РНК) и фосфатной группы. В молекуле ДНК присутствуют нуклеотиды с четырьмя видами оснований: аденин (обозначаемый как А), гуанин (G, Г), тимин (Т) и цитозин (С, Ц). В молекуле РНК вместо тимина присутствует урацил (U, У) (рис. 1.1, а). Связи между нуклеотидами в молекуле нуклеиновой кислоты образуются за счет (дезокси)рибозы и фосфатной группы. Таким

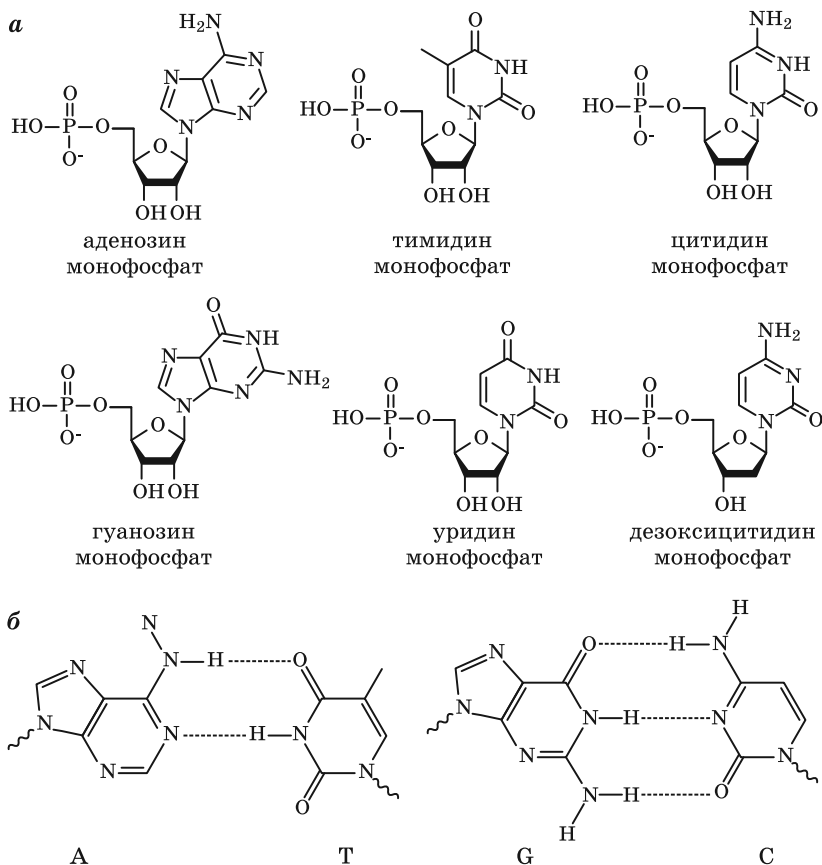


Рис. 1.1. Химическая структура нуклеотидов и дезоксинуклеотидов на примере дезоксцитидина (а) и формирование связей между пуринами и пиримидинами (б)

образом, каждая молекула нуклеиновой кислоты представляет собой последовательность нуклеотидов, различающихся между собой основаниями. Нуклеотидную последовательность молекулы РНК часто для простоты записывают как соответствующую ей молекулу ДНК, т. е. содержащую Т. Такой подход использован и в этой книге. Соответственно любая молекула нуклеиновой кислоты записывается как последовательность нуклеотидов с различными основаниями, например — АТАСГАГААААСГСА. Эта последовательность нуклеотидов является первичной структурой молекулы нуклеиновой кислоты. Нуклеотиды А и Г являются пуринами, а С и Т — пиримидинами. В двойной цепочке ДНК (вторичная структура), пурины А одной цепочки образуют две водородные связи с пиримидинами Т второй цепочки (слабая связь), а пурины Г — три связи с пиримидинами С (термодинамически более стабильная сильная связь) (рис. 1.1, б).

В молекуле нуклеиновой кислоты различают кодирующие и не кодирующие участки. Первые кодируют белки, вторые включают в себя, в частности, регуляторные элементы.

1.3. Аминокислотные последовательности

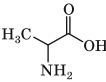
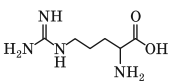

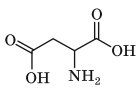
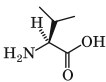
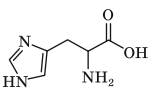

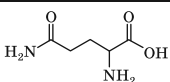
Основой структуры белка является полимерная цепочка, состоящая из мономеров-аминокислот — полипептид, аминокислотная последовательность. С химической точки зрения следует говорить: состоящая из аминокислотных остатков, однако в современной международной литературе по молекулярной эволюции для простоты принято использовать упрощенную терминологию.

Каждая аминокислота состоит из R-группы (боковой цепи), к которой присоединены карбоксильная и аминная группы. Связь между аминокислотами в полипептидной цепочке осуществляется при взаимодействии аминогруппы одной аминокислоты с карбоксильной группой другой аминокислоты (пептидная связь). Существует 20 так называемых стандартных аминокислот (табл. 1.1). Пользуясь различными критериями, аминокислоты можно отнести к разным типам или классам. Так, в зависимости от химического строения их боковых цепей выделяют алифатические (с дальнейшим делением на моноаминонокарбоновые, оксимоноаминокарбоновые и т. д.), ароматические и гетероциклические аминокислоты и иминокислоты. В молекулярной эволюции наиболее существенной является классификация аминокислот, основанная на их физико-химических свойствах: полярные или неполярные, нейтральные, кислые или

основные аминокислоты и т. д. Для обозначения аминокислот используются трехбуквенные или однобуквенные сокращения (см. табл. 1.1). Любая молекула белка может быть записана как последовательность аминокислот, например — Val-Val-Ile-Arg-Ser-Glu-Asn-Phe-Thr-Asp, или VVIRSENF^TD. Далее в нашей книге использованы однобуквенные сокращения.

Таблица 1.1

Химическая структура, обозначение и свойства аминокислот

Аминокислота	Химическая структура	Обозначение		Свойства
		трехбуквенное	однобуквенное	
1	2	3	4	5
аланин (alanine)		Ala	A	неполярная
аргинин (arginine)		Arg	R	положительно заряженная при pH = 7
аспарагин (asparagine)		Asn	N	полярная, незаряженная
аспарагиновая кислота (aspartic acid)		Asp	D	отрицательно заряженная при pH = 7
валин (valine)		Val	V	неполярная
гистидин (histidine)		His	H	положительно заряженная при pH = 7
глицин (glycine)		Gly	G	полярная, незаряженная
глутамин (glutamine)		Gln	Q	полярная, незаряженная

Продолжение табл. 1.1

1	2	3	4	5
глутаминовая кислота (glutamic acid)		Glu	E	отрицательно заряженная при pH = 7
изолейцин (isoleucine)		Ile	I	неполярная
лейцин (leucine)		Leu	L	неполярная
лизин (lysine)		Lys	K	положительно заряженная при pH = 7
метионин (methionine)		Met	M	неполярная
пролин (proline)		Pro	P	неполярная
серин (serine)		Ser	S	полярная, незаряженная
тирозин (tyrosine)		Tyr	Y	полярная, незаряженная
треонин (threonine)		Thr	T	полярная, незаряженная
триптофан (tryptophan)		Trp	W	неполярная
фенилаланин (phenylalanine)		Phe	F	неполярная
цистеин (cysteine)		Cys	C	полярная, незаряженная

1.4. Генетический код

Информация о порядке аминокислот в белковой цепочке содержится в кодирующих участках ДНК/РНК. Каждая аминокислота кодирована последовательностью из трех нуклеотидов —

Таблица 1.2

Универсальный генетический код

Прямая таблица		2-я позиция			
		Т	С	А	Г
1-я позиция	Т	ТТТ Phe/P	TCT Ser/S	TAT Tyr/Y	TGT Cys/C
		TTC Phe/P	TCC Ser/S	TAC Tyr/Y	TGC Cys/C
		TTA Leu/L	TCA Ser/S	TAA стоп	TGA стоп
		TTG Leu/L	TCG Ser/S	TAG стоп	TGG Trp/W
	С	СТТ Leu/L	CCT Pro/P	CAT His/H	CGT Arg/R
		СТС Leu/L	CCC Pro/P	CAC His/H	CGC Arg/R
		СТА Leu/L	CCA Pro/P	CAA Gln/Q	CGA Arg/R
		СТГ Leu/L	CCG Pro/P	CAG Gln/Q	CGG Arg/R
	А	АТТ Ile/I	ACT Thr/T	AAT Asn/N	AGT Ser/S
		АТС Ile/I	ACC Thr/T	AAC Asn/N	AGC Ser/S
		АТА Ile/I	ACA Thr/T	AAA Lys/K	AGA Arg/R
		АТГ Met/M	ACG Thr/T	AAG Lys/K	AGG Arg/R
	Г	ГТТ Val/V	GCT Ala/A	GAT Asp/D	GGT Gly/G
		ГТС Val/V	GCC Ala/A	GAC Asp/D	GGC Gly/G
		ГТА Val/V	GCA Ala/A	GAA Glu/E	GGA Gly/G
		ГТГ Val/V	GCG Ala/A	GAG Glu/E	GGG Gly/G

Обратная таблица			
Ala/A	GCA, GCT, GCG, GCC	Leu/L	TTA, TTG, CTA, CTT, CTG, CTC
Arg/R	AGA, AGG, CGA, CGT, CGG, CGC	Lys/K	AAA, AAG
Asn/N	AAT, AAC	Met/M	ATG
Asp/D	GAT, GAC	Phe/F	TTT, TTC
Cys/C	TGT, TGC	Pro/P	CCA, CCT, CCG, CCC
Gln/Q	CAA, CAG	Ser/S	AGT, AGC, TCA, TCT, TCG, TCC
Glu/E	GAA, GAG	Thr/T	ACA, ACT, ACG, ACC
Gly/G	GGA, GGT, GGG, GGC	Trp/W	TGG
His/H	CAT, CAC	Tyr/Y	TAT, TAC
Ile/I	ATA, ATT, ATC	Val/V	GTA, GTT, GTG, GTC
старт	ATG	стоп	TAA, TAG, TGA

триплетом, или кодоном. Соответствие между нуклеотидными триплетами и кодируемыми ими аминокислотами, а также сигналами начала и окончания синтеза полипептидной цепочки, трансляции, называют генетическим кодом. В подавляющем большинстве случаев генетический код в живых системах является одинаковым и называется универсальным, или стандартным, генетическим кодом (табл.1.2). Однако существуют и отличия от универсального генетического кода — как правило, небольшие, в значениях нескольких кодонов. Так, генетический код митохондрий имеет отличия от универсального кода. При этом митохондриальный генетический код одинаков у всех позвоночных, но отличается от митохондриальных генетических кодов, например, дрожжей или дрозофил (табл.1.3).

Поскольку кодон представляет собой последовательность из трех нуклеотидов и существуют четыре различных нуклеотида, то число возможных кодонов составляет $4^3 = 64$. При этом им необходимо кодировать 20 различных аминокислот и сигнал окончания трансляции, т. е. 21 сигнал (напомним, что отдельного сигнала для начала трансляции нет, стартовый кодон кодирует аминокислоту). Соответственно конкретные аминокислоты и сигнал окончания трансляции могут кодироваться более чем одним кодоном. Это явление называют вырожденностью генетического кода. Из 64 кодонов в универсальном генетическом коде 61 кодирует аминокислоты (смысловые кодоны), а три — сигнал окончания трансляции (нонсенс-, или стоп-кодоны). Различные кодоны, кодирующие одну и ту же аминокислоту, называют

Таблица 1.3

Примеры отличий от универсального генетического кода

Кодон	Значение в генетическом коде			
	универсальном	митохондриальном позвоночных	митохондриальном дрозофил	митохондриальном дрожжей
TGA	стоп-кодон	W	W	W
CTT	L	L	L	T
CTC	L	L	L	T
CTA	L	L	L	T
CTG	L	L	L	T
ATA	I	M	M	I
AGA	R	стоп-кодон	S	R
AGG	R	стоп-кодон	S	R

синонимичными кодонами. Так, изолейцин кодируется синонимичными кодонами АТА, АТТ и АТС. В большинстве случаев, синонимичные кодоны различаются между собой нуклеотидами в третьей позиции (см. табл. 1.2).

Отметим, что две аминокислоты, метионин и триптофан, кодируются только одним кодоном, а три аминокислоты — аргинин, лейцин и серин — кодируются шестью кодонами.

1.5. Мутации

В процессе репликации нуклеиновых кислот могут происходить ошибки. В результате дочерний геном будет отличаться от родительского генома. Ошибки, происходящие при репликации генома, называют мутациями (mutations). У многоклеточных организмов мутации могут происходить как в половых, так и в соматических клетках. Соматические мутации не передаются от поколения к поколению, и, таким образом, исключены из эволюционного процесса.

Мутации могут быть классифицированы в соответствии с числом затронутых ими нуклеотидов (длиной мутации) и типами происходящих при этом событий. Мутации могут затрагивать только один нуклеотид — точечные мутации (point mutations) — или несколько соседних нуклеотидов. При этом могут происходить следующие события: замена одного нуклеотида на другой — нуклеотидная замена (nucleotide substitution), вставка одного или более нуклеотидов (insertion), удаление одного или нескольких соседних нуклеотидов — делеция (deletion), поворот участка нуклеиновой кислоты длиной минимум в два нуклеотида на 180° — инверсия (inversion). Частным случаем вставки является удвоение некоего генетического участка — дупликация (duplication).

Особым случаем мутации является считывание дочерней молекулы нуклеиновой кислоты не с одной, а с двух и более родительских молекул — рекомбинация (recombination).

В кодирующих участках вставки и делеции, если их длина не кратна трем, приводят к изменению рамки считывания полипептидной цепочки. Это ведет к изменению всей последующей аминокислотной последовательности и, как правило, потере белком своей функции, нежизнеспособности такого потомства.

В молекулярной эволюции значительное внимание уделяется изучению нуклеотидных замен. Анализ рекомбинаций изложен в разд. 5.1.

1.6. Нуклеотидные замены

1.6.1. Транзиции и трансверсии

В кодирующих и некодирующих участках нуклеиновых кислот нуклеотидные замены могут быть разделены на транзиции (transitions) и трансверсии (transversions). Транзициями называют замены одного пурина на другой пурин ($A \rightarrow G$, $G \rightarrow A$) или одного пиримидина на другой пиримидин ($C \rightarrow T$, $T \rightarrow C$). Трансверсиями называют замены между пуринами и пиримидинами ($A \rightarrow T$, $A \rightarrow C$, $G \rightarrow T$, $G \rightarrow C$, $T \rightarrow A$, $T \rightarrow G$, $C \rightarrow A$, $C \rightarrow G$).

1.6.2. Синонимичные и несинонимичные замены

В кодирующих участках вследствие вырожденности генетического кода конкретная нуклеотидная замена может изменять или не изменять смысл кодона.

Нуклеотидную замену, не изменяющую кодируемую аминокислоту, называют синонимичной заменой (synonymous substitution, или silent substitution — молчащая). Так, нуклеотидная замена $A \rightarrow T$ в третьей позиции кодона АТА ($ATA \rightarrow ATT$) не изменяет кодируемую аминокислоту (до и после замены — изолейцин) и, таким образом, является синонимичной.

Нуклеотидную замену, изменяющую кодируемую аминокислоту, называют несинонимичной заменой (nonsynonymous substitution, или non-silent substitution — немолчащая). Так, замена $A \rightarrow G$ в третьей позиции того же кодона АТА ($ATA \rightarrow ATG$) является несинонимичной, поскольку кодируемая аминокислота изменяется с изолейцина на метионин. Таким образом, происходит замещение (replacement) одной аминокислоты другой (говорят о замене нуклеотида, но о замещении аминокислоты).

Мутацию, приводящую к изменению триплета с кодирующей аминокислоту на стоп-кодон, называют нонсенс-мутацией (nonsense mutation). Так, замена $TAC \rightarrow TAA$ является нонсенс-мутацией. Иногда нонсенс-мутации рассматривают как частный случай несинонимичных мутаций. В таком случае несинонимичные замены, приводящие к изменению одной аминокислоты на другую, но не на стоп-кодон, называют изменяющими смысл (missense mutations). Как правило, нонсенс-мутации приводят к потере кодируемым белком своей функции, нежизнеспособности потомства и соответственно исключаются из эволюционного процесса.

В любом кодоне каждый из трех нуклеотидов может быть заменен на любой из трех других нуклеотидов. Соответственно для каждого кодона число возможных замен составляет $3 \times 3 = 9$, т. е. при одной нуклеотидной замене каждый кодон может измениться на один из девяти других кодонов. Таким образом, для всех 64 кодонов число возможных замен составляет $64 \times 9 = 576$. Поскольку универсальный генетический код включает 61 смысловой кодон, то всего в смысловых кодонах возможны $61 \times 9 = 549$ нуклеотидных замен. В целом большинство этих замен будут несинонимичными. Однако, поскольку синонимичные кодоны, как правило, различаются нуклеотидами в третьей позиции, то большинство замен в третьей позиции будут синонимичными. Отметим, что некоторые замены в первой позиции кодона также являются синонимичными, но ни одна замена во второй позиции кодона синонимичной не является. Данные по возможным синонимичным и несинонимичным заменам для всех кодонов в каждой позиции приведены в табл. 1.4.

Таблица 1.4

**Возможные типы нуклеотидных замен
в кодирующих последовательностях**

Тип замены	Число замен	%
Всего во всех кодонах	549	100
синонимичные	134	25
несинонимичные	415	75
изменяющие смысл	392	71
нонсенс	23	4
Всего в первой позиции	183	100
синонимичные	8	4
несинонимичные	175	96
изменяющие смысл	166	91
нонсенс	9	5
Всего во второй позиции	183	100
синонимичные	0	0
несинонимичные	183	100
изменяющие смысл	176	96
нонсенс	7	4
Всего в третьей позиции	183	100
синонимичные	126	69
несинонимичные	57	31
изменяющие смысл	50	27
нонсенс	7	4

Говорят и о синонимичных и несинонимичных позициях кодона. Если все три возможные замены нуклеотида в данной позиции не изменяют кодируемую аминокислоту, то такая позиция является синонимичной. Если две из трех возможных замен нуклеотида в данной позиции не изменяют кодируемую аминокислоту, то такая позиция является синонимичной на две трети, и т. д.

1.7. Нуклеотидный и аминокислотный состав, использование кодонов

Процентное содержание различных нуклеотидов и аминокислот в генетической последовательности называют ее нуклеотидным и аминокислотным составом, или композицией (nucleotide composition, amino acid composition). Если замена любого нуклеотида на любой другой нуклеотид при репликации последовательности, а также последующее закрепление любой нуклеотидной замены происходят с одинаковой вероятностью, то содержание любого из четырех нуклеотидов в последовательности будет приблизительно равно 25%. Однако содержание конкретных нуклеотидов в геномах может значительно отклоняться от 25%. В таких случаях говорят о смещениях нуклеотидного состава (nucleotide bias) или предпочтении того или иного нуклеотида (nucleotide preference). Скажем, если в некоей последовательности содержание нуклеотидов А, Т, G и С составляет 35, 15, 22 и 28% соответственно, то говорят о преобладании нуклеотида А за счет нуклеотида Т в составе этой последовательности. Рассматривают также суммарное содержание нуклеотидов G и С — GC-содержание (GC-content). Поскольку в двойной цепочке ДНК нуклеотиды G и С образуют между собой три связи (разд. 1.2), то двухцепочечная молекула ДНК с большим GC-содержанием термодинамически более стабильна (ее температура плавления выше).

Как уже отмечалось (разд. 1.4, см. табл. 1.2), вследствие вырожденности генетического кода, большинство аминокислот может кодироваться более чем одним кодоном. Показано, что в кодирующих последовательностях могут наблюдаться предпочтения при выборе тех или иных синонимичных кодонов [Grantham *et al.*, 1980]. Скажем, лейцин кодируется шестью кодонами, однако в наружном мембранном белке II (*ompA*) *Escherichia coli* для кодирования лейцина в 21 из 23 случаев

используется лишь один кодон — СТГ. При описании распределения различных синонимичных кодонов в последовательности говорят об использовании кодонов (codon usage).

Вопросы эволюционного анализа смещения нуклеотидного состава и использования кодонов обсуждены в разд. 5.2.

1.8. Эволюция нуклеотидной последовательности

Рассмотрим эволюцию молекулы нуклеиновой кислоты. Пусть в момент времени 0 эта молекула длиной в 30 нуклеотидов имеет следующую последовательность 0:

```

                                111111111122222222223
позиция                        123456789012345678901234567890
последовательность 0 ATACGAGAAAACGCAATCCGCGAGAATGCC

```

При репликации этой молекулы могут происходить или не происходить ошибки. Если за некий промежуток времени (измеряемый либо единицами времени, либо числом репликаций последовательности, т. е. числом поколений) при репликации этой молекулы ошибки происходить не будут, то существующая в момент времени 1 дочерняя последовательность 1 не будет отличаться от родительской последовательности 0. Соответственно можно говорить, что за промежуток времени 0–1 рассматриваемая последовательность не претерпела эволюционных изменений. Для наглядности эти две последовательности можно записать одну под другой, а нуклеотиды последовательности 1, совпадающие с таковыми в той же позиции последовательности 0, обозначить знаками тире:

```

                                111111111122222222223
                                123456789012345678901234567890
0   ATACGAGAAAACGCAATCCGCGAGAATGCC
1   -----

```

Предположим, что при дальнейшей репликации этой молекулы будут происходить ошибки. Так, пусть за промежуток времени 1–2 в рассматриваемой последовательности произошли две нуклеотидные замены: замена С → А в позиции 4 и замена С → Т в позиции 21. В таком случае на момент времени 2 порядок эволюционных изменений в рассматриваемой последовательности можно записать следующим образом:

```

1111111111222222222223
123456789012345678901234567890
0  ATACGAGAAAACGCAATCCGCGAGAATGCC
1  -----
2  ---A-----T-----

```

Последовательности 0 и 1 не отличаются друг от друга, являются идентичными (identical). Последовательности же 1 и 2 являются не идентичными, но похожими (similar). Они имеют различия между собой, иначе говоря — они находятся на определенной генетической, или эволюционной, дистанции (genetic distance, evolutionary distance), или расстоянии, друг от друга.

Пусть за последующий промежуток времени 2–3 в рассматриваемой последовательности произошли еще три замены: в позициях 4 ($A \rightarrow C$), 21 ($T \rightarrow A$) и 30 ($C \rightarrow G$):

```

1111111111222222222223
123456789012345678901234567890
0  ATACGAGAAAACGCAATCCGCGAGAATGCC
1  -----
2  ---A-----T-----
3  -----A-----G

```

При этом замена нуклеотида в позиции 30 произошла впервые (первичная замена), в то время как в позициях 4 и 21 нуклеотидные замены ранее уже происходили — в этих позициях за промежуток времени 2–3 произошли вторичные замены. Таким образом, за весь рассмотренный промежуток времени 1–3 в позициях 4 и 21 произошли множественные нуклеотидные замены (multiple hits at one site, множественные удары в одной позиции): в позиции 4 произошли множественные замены $C \rightarrow A \rightarrow C$, в позиции 21 — множественные замены $C \rightarrow T \rightarrow A$.

Как видно, вторичная замена в позиции 4 привела к тому, что нуклеотид, находящийся в этой позиции, снова совпадает с таковым у предка, последовательности 0. Такие вторичные замены называют обратными заменами (back substitutions), или реверсиями (reversions).

Пусть в дальнейшем за промежуток времени 3–4 в рассматриваемой последовательности произошла делеция нуклеотидов 13–15 и замена нуклеотида в позиции 29 ($C \rightarrow A$), а затем,

за промежуток времени 4–5 — замена нуклеотида в позиции 18 (С → Т). Обозначая делеции точками (иногда используют и противоположные обозначения: совпадающие нуклеотиды — точки, делеции — тире), все эти эволюционные изменения, т. е. эволюционную историю рассматриваемых последовательностей, можно записать следующим образом:

```

                                111111111122222222223
                                123456789012345678901234567890
0   ATACGAGAAAAACGCAATCCGCGAGAATGCC
1   -----
2   ---A-----T-----
3   -----A-----G
4   -----...-A-----AG
5   -----...-T--A-----AG

```

Позиции, в которых в эволюционной истории происходило множество замен, называют изменчивыми, или вариабельными (variable), позициями. Так, в рассматриваемом случае позиции 4 и 21 — изменчивые. Позиции, в которых замены не происходили, или их было мало, называют соответственно абсолютно или относительно консервативными (conserved). Абсолютно консервативные позиции называют также инвариантными. Так, позиция 1 — абсолютно консервативная, а позиции 18, 29 и 30 — относительно консервативные. Говорят и об изменчивых и консервативных участках последовательности. Так, в рассматриваемом случае участок 1–12 или 1–20 можно назвать относительно консервативным, а участок 21–30 — относительно изменчивым.

В рассматриваемом случае мы исходили из того, что каждая родительская последовательность дает только одну дочернюю последовательность. Естественно, в общем случае это не так. Каждая родительская последовательность может быть реплицирована несколько раз, и происходящие при каждом цикле ее репликации ошибки могут быть различны. В таком случае дочерние последовательности одной родительской последовательности будут различаться между собой. Более того, даже если дочерние последовательности одинаковы, ошибки при их дальнейшей репликации будут, в общем случае, различны.

Рассмотрим такой случай (рис. 1.2). Пусть в момент времени 0 молекула нуклеиновой кислоты имеет последовательность 0. Предположим, что за каждый промежуток времени одна роди-

тельская последовательность дает две дочерние последовательности. Пусть за промежуток времени 0–1 последовательность 0 даст две не отличающиеся друг от друга идентичные дочерние последовательности 1.1 и 1.2. Очевидно, что при последующей репликации мутации, возникающие в последовательностях 1.1 и 1.2, будут, в общем случае, различными: мутации при репликации последовательности 1.1 не зависят от того, какие мутации произошли при репликации последовательности 1.2. Если между потомками этих последовательностей (форм жизни, от которых они получены) не будет происходить обмена генетической информацией (отсутствие полового процесса, рекомбинаций и т. д., не важно, в силу каких именно причин), то дальнейшая эволюция последовательности 1.1 не будет зависеть от эволюции последовательности 1.2. В таком случае говорят, что при эволюции последовательности 0 в поколении (момент времени) 1 произошла дивергенция между дочерними линиями (divergence), разделение на две независимые родственные эволюционные линии (evolutionary lineages), возможное начало видообразования. Говорят также об ответвлении (branching out) эволюционных линий, их почковании, разветвлении эволюционного дерева. Отметим, что в последующем в дочерних последовательностях этих линий могут происходить и одинаковые мутации: так, в рассматриваемом примере в этих эволюционных линиях независимо друг от друга произошли одинаковые мутации в позиции 18 ($A \rightarrow T$, рис. 1.2). Одинаковые мутации, которые произошли в родственных эволюционных линиях независимо друг от друга, называют параллельными мутациями (parallel mutations).

Таким образом, через промежуток времени 0–3 потомство последовательности 0 будет представлять собой восемь различающихся между собой последовательностей — генетически гетерогенную (heterogeneous) группу последовательностей. Процесс, ведущий к образованию такой гетерогенной группы, называют диверсификацией (diversification). Все рассматриваемые на рис. 1.2 последовательности происходят от единого общего предка — последовательности 0. Последовательности, имеющие общее эволюционное происхождение, называют гомологичными (homologous) (разд. 1.10). Группу таких гомологичных последовательностей называют монофилетической (monophyletic) группой, имеющей общее эволюционное происхождение, общего предка. Поскольку задачей молекулярной филогенетики является установление эволюционной истории генов и форм жизни после их

дивергенции от общего предка, то эта область молекулярной эволюции по определению анализирует гомологичные последовательности. Выяснение вопроса, являются ли некие последовательности гомологичными, а не просто похожими друг на друга (например, в результате конвергентной эволюции, разд.1.10), является важной задачей эволюционного анализа.

Одним из ключевых понятий молекулярной эволюции является понятие последнего общего предка группы последовательностей (в дословном переводе с английского — наиболее недавнего общего предка, *most recent common ancestor*, MRCA) и времени его существования. Последний общий предок всех форм жизни на Земле обозначается термином LUCA (*last universal common ancestor*). Для всех рассматриваемых последовательностей поколения 3 последний общий предок существовал в поколении (момент времени) 1, а не в поколении 0, поскольку диверсификация группы последовательностей поколения 3 произошла после момента времени 1.

При рассмотрении представленного на рис. 1.2 примера эволюции нуклеотидной последовательности мы исходили из того, что для поколения 3 известны все предки, родительские последовательности, порядок диверсификации последовательностей в группе, все мутации, происходившие в их эволюционной истории. Естественно, что на практике — ни вымершие (*extinct*) предки неких существующих сегодня последовательностей, ни порядок их диверсификации, ни происходившие при эволюции последовательностей мутации неизвестны. Собственно, задачей молекулярной эволюции и является установление сценария событий, имевших место в эволюционной истории анализируемой группы последовательностей, форм жизни. При этом базовый принцип молекулярной эволюции как науки основан на положении о том, что наиболее вероятный, наиболее близкий к истинному сценарий эволюции группы последовательностей — это тот сценарий, который требует минимального числа эволюционных событий. Безусловно, это может быть и не так, и истинным — может быть и другой сценарий, с бóльшим числом событий, но в данном случае действует научный принцип, известный как бритва, или лезвие, Оккама: без необходимости не следует утверждать многое (лат. — *Pluralitas non est ponenda sine necessitate*). То, что можно объяснить посредством меньшего, не следует выражать посредством большего. В науке под бритвой Оккама понимают общий принцип, утверждающий, что если существует несколько логически непротиворечивых определе-

ний или объяснений какого-либо явления, то верным следует считать наиболее простое из них.

Итак, задачей молекулярной эволюции является установление эволюционной истории группы гомологичных последовательностей, генов, организмов, живых систем вообще — например, эволюционной истории существующих сегодня восьми последовательностей поколения 3, о которых имеется генетическая информация (см. рис. 1.2). Запишем эту информацию, учитывая, что предки этих последовательностей нам неизвестны, и неизвестно, в каких участках каких именно современных последовательностей имели место делеции или вставки:

```
1 ATCAGGGAGAATAAAGCT
2 ATGCGAGAGCCTAGT
3 ATTCGAGAGAATAAAGAA
4 ATCCGTGAGAATAAAGCG
5 ATCAGGGAGAATAAAGCC
6 ATGCGAGAGAATGCT
7 ATTCGAGAGAATAAAGAA
8 ATCCGTGAGAATAAAGCG
```

Первым этапом филогенетического анализа группы последовательностей является идентификация вставок и делеций, имевших место в их эволюционной истории — выравнивание последовательностей (гл. 2). Методы эволюционного анализа последовательностей изложены в гл. 3–5.

1.9. Консенсусные последовательности

Для анализируемой группы последовательностей можно рассмотреть так называемую консенсусную последовательность, или консенсус (consensus) — искусственную последовательность, в каждой позиции которой находится нуклеотид, наиболее часто встречающийся у анализируемых последовательностей. Как правило, при расчете консенсусной последовательности в нее записывают нуклеотид, присутствующий как минимум в 50% анализируемых последовательностей (консенсус-50%). При этом отсутствие нуклеотида в конкретной позиции (одной из) анализируемых последовательностей можно принимать или не принимать во внимание. Если в 50% последовательностей в данной позиции присутствует один нуклеотид, а в других 50% последовательностей — другой нуклеотид, то в консенсусную последовательность

принято записывать нуклеотид из первой последовательности группы. При расчете консенсусной последовательности можно использовать и менее, и более строгие пороговые критерии. Так, в консенсус можно включать нуклеотиды, просто наиболее часто встречающиеся в конкретной позиции, даже если они присутствуют менее чем в 50% последовательностей, или наоборот — минимум в 90% последовательностей, или даже в 100% последовательностей. Если гетерогенность конкретной позиции не отвечает заданным критериям — скажем, выбран порог в 50%, но ни один из нуклеотидов не присутствует минимум в 50% последовательностей — то такую позицию принято обозначать знаком вопроса, или N, или X. Все эти правила действительны и для аминокислотных последовательностей.

Кроме того, в консенсусной нуклеотидной последовательности можно отражать разнообразие нуклеотидов в конкретной позиции (гетерогенность позиции, *heterogeneity*). Для этого используют коды Международного союза теоретической и прикладной химии IUPAC (International Union of Pure and Applied Chemistry). Так, если в некоей конкретной позиции в одних последовательностях находится А, а у других — Т, то А/Т гетерогенность этой позиции в консенсусной последовательности можно отразить как W. Коды IUPAC для обозначения нуклеотидов приведены в табл. 1.5.

Для рассмотренного выше примера (пусть нам известны только последовательности поколения 3 на рис. 1.2) консенсусные последовательности можно записать следующим образом:

	111111111
	123456789012345678
консенсус-100%	AT??G?GAG??T???G??
консенсус-50%	ATCCGAGAGAATAAAGC?
консенсус-IUPAC	ATBMGDGAGMMTRVWGMN
1	ATCAGGGAGAATAAAGCT
2	ATGCGAGAGCCTAGT
3	ATTTCGAGAGAATAAAGAA
4	ATCCGTGAGAATAAAGCG
5	ATCAGGGAGAATAAAGCC
6	ATGCGAGAGAATGCT
7	ATTTCGAGAGAATAAAGAA
8	ATCCGTGAGAATAAAGCG

Таблица 1.5

Коды IUPAC для обозначения нуклеотидов

Код	Обозначает	Комплементарный нуклеотид
A	A	T (U)
C	C	G
G	G	C
T (U)	T (U)	A
M	A или C	K
R	A или G	Y
W	A или T (U)	W
S	C или G	S
Y	C или T (U)	R
K	G или T (U)	M
V	A или C или G	B
H	A или C или T (U)	D
D	A или G или T (U)	H
B	C или G или T (U)	V
X или N	A или C или G или T (U)	X или N

Для наглядности индивидуальные последовательности можно записать относительно консенсуса (скажем, 50%-го), используя тире для обозначения одинаковых нуклеотидов:

```

                                111111111
                                123456789012345678
консенсус-50%                 ATCCGAGAGAATAAAGC?
1                               ---A-G-----T
2                               --G-----CC--GT
3                               --T-----AA
4                               ----T-----G
5                               ---A-G-----C
6                               --G-----GCT
7                               --T-----AA
8                               ----T-----G

```

Поскольку последовательности поколения 3 имеют разную длину, то в их эволюционной истории происходили вставки и/или делеции нуклеотидов. Однако, поскольку эволюционная

история поколения 3 нам (как мы договорились) неизвестна, то неизвестно, где именно произошли вставки-делеции. Эти вопросы устанавливаются в процессе выравнивания последовательностей (гл. 2).

1.10. Гомологичные и сходные признаки, конвергенция

В разд. 1.8 подчеркивалось, что молекулярная филогенетика анализирует гомологичные последовательности (имеющие общее эволюционное происхождение). Наличие у нескольких организмов похожих признаков не обязательно является свидетельством общего эволюционного происхождения этих признаков и соответственно самих организмов. Классическим примером сходного, или подобного, признака, имеющего независимое эволюционное происхождение, является наличие крыльев у насекомых, птиц и летучих мышей. Наличие таких сходных признаков называют гомоплазией (homoplasy), а сами признаки — не гомологичными, но аналогичными (analogous). Причиной гомоплазии является конвергентная эволюция (convergent evolution, сходящаяся), конвергенция — эволюционный процесс, при котором у неродственных организмов в процессе адаптации к одним и тем же природным условиям независимо образуется сходный признак.

Наличие того или иного нуклеотида (аминокислоты) в определенной позиции последовательности также является призна-

Таблица 1.6

Дивергентная, параллельная и конвергентная эволюция

	Эволюция						
	дивергентная		параллельная		конвергентная		
Родительские последовательности	1	2	1	2	1	2	
Нуклеотиды в позиции i у родительских последовательностей	A		A	A	A		T
Нуклеотиды в той же позиции у дочерних последовательностей	A	T	C	T	T	A	G

ком. Вышеизложенные рассуждения важны для понимания того, что схожесть между некоторыми генетическими последовательностями, наличие у них одинаковых нуклеотидов (аминокислот) в соответствующих позициях не является достаточным основанием для заключения о гомологичности этих позиций, последовательностей вообще. Наличие одного и того же нуклеотида (аминокислоты) может быть и результатом конвергентной эволюции. Примеры дивергенции, конвергенции и параллельной эволюции приведены в табл. 1.6. Отметим, что на молекулярном уровне конвергенция белковых молекул не сводится к схожести их аминокислотных последовательностей: функциональное сходство белков может достигаться и схожестью их архитектуры.

1.11. Естественный отбор и неодарвинизм

В рассматриваемых в разд. 1.8 примерах мы исходили из того, что любые мутации в родительских геномах не изменяют приспособленности (*fitness*, *adaptation*) дочерних геномов к окружающей среде — характеристик, сводящихся в конечном счете к плодовитости родителей и их потомства. Соответственно мы подразумевали, что все мутации имеют одинаковые, 100%-ные, вероятности быть переданными дочерним геномам и закрепиться в популяции потомков. Однако сформулированная Дарвином теория естественного отбора исходит из того, что вероятность закрепления (отбора) некоего наследственного признака в популяции зависит от качеств этого признака — от того, насколько он способствует преимущественному выживанию и преимущественному производству плодovитого потомства особью, этим признаком обладающей [*Darwin*, 1871]. Теория естественного отбора, законы наследственности и динамика генов в популяции подробно изучаются в курсах общей биологии и генетики. Если у особи в популяции произошла мутация, и через некое время эта мутация присутствует у всех особей этой популяции, то говорят о закреплении, фиксации (*fixation*) данной мутации в популяции. Если через некое время эта мутация отсутствует у всех особей этой популяции, то говорят о потере (*loss*) или вымирании (*extinction*) данной мутации в популяции.

В современной науке эволюционные идеи Дарвина объединены с данными других наук, прежде всего — генетики, в синтетическую теорию эволюции, называемую неодарвинизмом. В молекулярной эволюции неодарвинизм понимает мутацию как единственный источник изменчивости, а факторы естественного

отбора — как ведущие движущие силы эволюции (evolutionary driving forces), определяющую судьбу мутаций — их преимущественное закрепление или потерю. Неодарвинизм подразумевает, что конкретная мутация может изменить — улучшить или ухудшить — или не изменить приспособленность организма к окружающей среде. Мутации, улучшающие приспособленность организма, подвергаются действию положительного естественного отбора — эволюционным силам, факторам, направленным на отбор такой мутации, ее преимущественное закрепление в популяции. Мутации, ухудшающие приспособленность организма, подвергаются действию отрицательного естественного отбора — эволюционным силам, факторам, направленным на преимущественное удаление такой мутации из популяции. Мутации, не изменяющие приспособленности организма к окружающей среде, называют селективно нейтральными (neutral). Вероятность их закрепления или удаления из популяции определяется не эволюционными силами, а исключительно вероятностными, стохастическими процессами. Процесс изменения частоты мутации в популяции под действием стохастических процессов называют случайным генетическим дрейфом (random genetic drift).

Для кодирующих нуклеотидных последовательностей основными факторами естественного отбора являются эволюционные факторы, действующие на уровне белка. Соответственно действию этих факторов подвергаются несинонимичные, изменяющие белок, но не синонимичные замены. В целом синонимичные замены принято считать селективно нейтральными, или близкими к нейтральным. Однако существуют и факторы естественного отбора, действующие преимущественно и даже исключительно на уровне синонимичных замен (разд. 5.2).

1.12. Закрепление мутации в популяции

Математически приспособленность некой исходной последовательности (дикого типа, wild type) к окружающей среде можно обозначить за 1, а изменение ее приспособленности после некой мутации, селективное преимущество или селективный недостаток у дочерней, мутантной последовательности — за s . В таком случае приспособленность дочерней последовательности к окружающей среде будет равна $1 + s$, где s имеет положительное или отрицательное значения для мутаций, улучшающих или ухудшающих приспособленность последовательности, соответственно. Вероятность закрепления (отбора) этой мутации в популяции, P ,

будет зависеть от s . Теория естественного отбора подразумевает преимущественное, но не обязательное закрепление либо удаление мутации, соответственно улучшающей («полезная» мутация) или ухудшающей («вредная» мутация) приспособленность. Таким образом, даже для мутации, ухудшающей приспособленность последовательности, существует ненулевая вероятность быть закрепленной в популяции.

При обсуждении вероятности отбора мутации необходимо ввести понятие эффективного размера, или численности, популяции (effective population size), N_e . В популяционной биологии размер популяции N определяется как общее число особей (последовательностей) в популяции. Однако (далеко) не все они принимают участие в репродукции, соответственно — в эволюционных процессах. В первом приближении эффективный размер популяции можно понимать как ту часть популяции, которая принимает участие в размножении, соответственно — в эволюционных процессах. Более строго понятие эффективного размера популяции определяется как число размножающихся особей, достаточное для описания процессов распределения частот мутаций во всей популяции при случайном генетическом дрейфе [Wright, 1931; Wright, 1938].

Показано, что вероятность закрепления мутации в популяции диплоидных организмов определяется формулой

$$P = \frac{1 - e^{-4N_e s q}}{1 - e^{-4N_e s}},$$

где q — начальная частота мутации в популяции [Kimura, 1962].

Если в популяции размером N происходит одна мутация, то ее начальная частота для диплоидных организмов будет равна $(1/2)N$.

Поскольку $e^{-x} \approx (1 - x)$ при малых x , то при приближении s к 0 (т. е. для мутаций, близких к нейтральным) $P \approx q$. Таким образом, вероятность нейтральной мутации быть зафиксированной в популяции равна ее начальной частоте. Соответственно, чем меньше размер популяции, тем более вероятно закрепление нейтральной мутации (эффект бутылочного горлышка, см. ниже).

Если эффективный размер популяции равен размеру популяции, уравнение для расчета вероятности закрепления мутации в популяции диплоидных организмов можно переписать как

$$P = \frac{1 - e^{-2s}}{1 - e^{-4Ns}}.$$

При небольших положительных значениях s (разумно считать, что каждая отдельная мутация будет лишь незначительно увеличивать приспособленность особи) и больших значениях N , $P \approx 2s$. Таким образом, вероятность фиксации мутации с $s = 0,01$ составит 2%.

Рассмотрим численный пример. Предположим, что в популяции, состоящей из тысячи особей, произошли мутации с $s = 0,01$ и $-0,001$. Если упрощенно предполагать, что $N_e = N$, то вероятности закрепления этих мутаций будут составлять 0,05, 2 и 0,004% соответственно. Таким образом, даже улучшающие приспособленность организма к окружающей среде мутации с $s = 0,01$ будут в 98% случаев, тем не менее, потеряны, удалены из популяции в силу случайных событий. Даже при наличии факторов естественного отбора, случайные события играют значительную роль в эволюции живых систем.

Отметим, что роль случайных событий возрастает при уменьшении N_e . Возрастание роли случайных событий при резком уменьшении численности популяции называют эффектом генетической воронки, или бутылочного горлышка (bottleneck effect). В предельном случае, когда от некой гетерогенной популяции (для простоты будем рассматривать гаплоидные бесполое организмы) остается только одна особь — например, в результате катастрофы, — то все имеющиеся у этой особи мутации, включая и ухудшающие приспособленность, будут со 100%-ной вероятностью закреплены в дочерней популяции. В таких случаях говорят об эффекте основателя (founder effect) [Mayr, 1954; Mayr, 1963]. Подобные эффекты проявляются и в случае проникновения ограниченного числа особей в новую восприимчивую среду обитания, что часто происходит, например, при распространении патогенов. В таких случаях победитель — имеется в виду первый проникший в новую среду обитания — получает все, т.е. закрепляет в дочерней популяции все имеющиеся у него мутации. Эффект основателя наблюдается и при межвидовом переносе вирусов — скажем, парвовируса кошкам собакам [Lukashov and Goudsmit, 2001], — и при внутривидовом распространении вируса в новую, ранее этим вирусом не пораженную, популяцию — скажем, при распространении вируса иммунодефицита человека первого типа, ВИЧ-1, на территории бывшего СССР (подробно рассмотрено в разд. 5.5.3).

Кроме собственно вероятности закрепления мутации в популяции, важно рассмотреть и время закрепления мутации, t — число поколений, необходимое для закрепления мутации. При

этом следует говорить о среднем условном времени закрепления мутации, учитывая тот факт, что большинство мутаций будет потеряно.

Показано, что если начальная частота мутации в популяции диплоидных организмов составляет $(1/2)N$, то для нейтральных мутаций, которые в конечном счете будут закреплены в популяции, среднее условное время закрепления t составит $4N$ поколений [Kimura and Ohta, 1969].

Для мутаций с селективным преимуществом s , которые будут закреплены в популяции, среднее условное время закрепления составит

$$t = \frac{2}{s} \ln(2N) \text{ поколений.}$$

Таким образом, для популяции диплоидных организмов численностью в миллион особей со временем репродукции в два года, для закрепления нейтральной мутации потребуется в среднем 8 миллионов лет. Для сравнения, закрепленные в конечном счете в этой популяции мутации с селективным преимуществом в 0,01 будут в среднем закреплены за 5800 лет.

Показано, что для мутации с селективным недостатком s среднее условное время закрепления равно среднему условному времени закрепления мутации с селективным преимуществом s [Maruyama and Kimura, 1974]. Таким образом, закрепление мутации с селективным недостатком в 0,01 также произойдет в среднем за 5800 лет. Этот, на первый взгляд, парадоксальный математический результат, тем не менее, легко понять с точки зрения естественного отбора. Движущие силы естественного отбора направлены на преимущественное удаление из популяции мутации с селективным недостатком. Однако судьба такой мутации определяется и стохастическими, случайными процессами. Если стохастические процессы в течение достаточно короткого времени не приведут случайным образом к закреплению такой мутации, то она будет удалена из популяции постоянно действующими факторами отрицательного отбора. Соответственно такая мутация будет либо закреплена в популяции относительно быстро, либо, в противном случае, потеряна. Таким образом, улучшающие и ухудшающие приспособленность организма мутации с одинаковым по модулю значением s будут в среднем условно закреплены за одинаковое время. Однако процент улучшающих приспособленность мутаций, закрепленных в популяции, будет значительно большим.

1.13. Концепция молекулярных часов

Поскольку ошибки, происходящие при репликации генетического материала, могут происходить при каждом цикле репликации, то, чем больше проходит этих циклов, тем больше может произойти ошибок, мутаций. Кроме того, закрепление мутации в популяции занимает определенное время. Таким образом, чем больше времени проходит после момента дивергенции последовательностей, форм жизни, тем, в общем случае, больше эволюционных различий будет наблюдаться между ними.

Впервые этот феномен был отмечен при анализе аминокислотных последовательностей гемоглобина и цитохрома С у различных организмов [Zuckermandl and Pauling, 1962; Margoliash, 1963]. В дальнейшем Цукеркандль и Полинг сформулировали концепцию существования молекулярных часов (molecular clock) в эволюции живых систем [Zuckermandl and Pauling, 1965]. Эта концепция утверждает, что для конкретной генетической последовательности скорость эволюции постоянна во времени и одинакова у всех дочерних последовательностей, т.е. во всех нисходящих эволюционных линиях. Иначе говоря, если известно, что дивергенция двух различающихся между собой на один нуклеотид последовательностей произошла 10 лет тому назад, то дивергенция между этими двумя последовательностями и третьей последовательностью, имеющей два отличия от них, произошла 20 лет тому назад (доверительный интервал для этой даты может быть рассчитан с помощью соответствующего математического аппарата).

Математические подходы, основанные на концепции молекулярных часов, широко используются при изучении филогенетических отношений между формами жизни и датировании эволюционных событий. Вопросы, связанные с анализом молекулярных часов, подробно разбираются в разд. 5.3.

1.14. Нейтральная теория молекулярной эволюции

В конце 1960-х гг. Кимура [Kimura, 1968] и независимо от него — Кинг и Джукс [King and Jukes, 1969] сформулировали гипотезу, получившую впоследствии название нейтральной теории молекулярной эволюции (neutral theory of molecular evolution) [Kimura, 1983]. Книга Кимуры переведена на русский язык [Кимура, 1985]. В отличие от (традиционно понимаемого) дарвинизма, нейтральная теория молекулярной эволюции посту-

лирует, что на молекулярном уровне подавляющее большинство эволюционных изменений (практически — все изменения) определяются не действием факторов положительного отбора, а случайным генетическим дрейфом селективно нейтральных (или близких к таковым) мутаций. Согласно нейтральной теории, судьба мутаций в популяции определяется исключительно стохастическими процессами. Генетическая гетерогенность любой популяции, таким образом, в любой момент времени представляет собой переходное состояние, некий разрез динамического процесса, в котором каждая конкретная мутация находится на пути либо к ее закреплению в популяции, либо к ее потере. Соответственно, с точки зрения нейтральной теории молекулярной эволюции, любая конкретная наблюдаемая генетическая гетерогенность популяции по своей сути преходяща.

Теория нейтральной эволюции в значительной степени базируется на экспериментальных данных о молекулярных часах в эволюции живых систем (хотя, по мнению автора этой книги, существование молекулярных часов можно понять и принять и с точки зрения дарвинизма).

Теория нейтральной эволюции вызвала (и продолжает вызывать) значительную критику со стороны (традиционных) дарвинистов, но одновременно явилась катализатором становления и развития многих идей и методов, лежащих в основе современного эволюционного анализа. Если эволюция (в значительной степени) определяется стохастическими процессами, то для описания, анализа эволюционного процесса должен и может быть применен имеющийся математический аппарат, описывающий случайные процессы, теория вероятности.

До сегодняшнего времени нейтральную теорию молекулярной эволюции принято противопоставлять неodarвинизму. Пример дискуссии между этими направлениями обсужден в разд. 5.2.1. Однако, по мнению автора этой книги, эти два направления эволюционного учения можно и следует понимать не как противостоящие, взаимоисключающие, но как дополняющие, обогащающие друг друга. Обе теории исходят из того, что большинство новых мутаций ухудшают приспособленность организма, быстро теряются в популяции и, таким образом, не наблюдаются при эволюционном анализе. Далее, неodarвинизм постулирует обязательного отбора мутаций, улучшающих приспособленность организма, но говорит о преимущественном их закреплении в популяции. Как обсуждалось в разд. 1.12, улучшающие приспособленность организма к окружающей среде мутации могут

быть, тем не менее (и даже — в подавляющем большинстве случаев), потеряны. Таким образом, их судьба с точки зрения и неodarвинизма, и теории нейтральной эволюции, в значительной степени определяется стохастическими процессами. С другой стороны, нейтральная теория не отрицает наличия мутаций, улучшающих приспособленность организма, не утверждает, что абсолютно все закрепленные мутации являются нейтральными, но постулирует, что таких мутаций — (подавляющее) большинство.

1.15. Эволюционная систематика

Как отмечалось в разд. 1.1, одной из ключевых задач молекулярной эволюции как науки является создание классификации форм жизни — систематики, таксономии.

Безусловно, понимание необходимости и стремление решить эту задачу существовали задолго до начала молекулярной эры в биологии. Еще Линнеем были сформулированы принципы систематики растений, живых (и неживых — минералы) систем вообще [*Linnaei*, 1735]. Не останавливаясь подробно на анализе предложений Линнея, описанных в имеющейся учебной литературе, отметим ключевое для нашей книги положение: при разработке классификации форм жизни целью Линнея было исключительно облегчение идентификации растений и животных как представителей либо нового, либо уже известного вида, рода, семейства и т. д. После изучения морфологических признаков некоего организма, он мог быть легко классифицирован путем сравнения выявленных признаков с описанными для известных видов. Принципы систематики Линнея, ее последующее развитие базировались исключительно на установлении схожести наблюдаемых морфологических признаков форм жизни, но не на установлении их общего эволюционного происхождения. Одним из классических примеров здесь является использование наличия у животных копыт как важного классифицирующего признака, в связи с чем выделялась группа копытных животных (надотряд *Ungulata*), объединявшая в том числе парнокопытных коров (отряд *Artiodactyla*) и непарнокопытных лошадей (отряд *Perissodactyla*). Наличие копыт у некоего нового вида животных позволяло классифицировать этот вид как представителя группы копытных с последующей более детальной классификацией на основании других морфологических признаков (включая и число копыт).

Предложенные Линнеем подходы к систематике не базировались, и не могли базироваться, на принципе общего эволюционного происхождения форм жизни. Доминировавшая во времена Линнея догма о неизменности видов, полностью им разделяемая, делала невозможным саму постановку вопроса об эволюционных отношениях между формами жизни, их филогении.

По мере развития палеонтологии, показавшей изменение форм жизни, и эволюционного учения вообще, линнеевские принципы систематики смогли достаточно просто (бесконфликтно и быстро, в отличие от ряда других направлений биологии) учесть идеи об изменчивости видов. Принципы Линнея стали применяться и для описания ископаемых останков, окаменелостей (fossils), систематики вымерших организмов как представителей той или иной группы существующих организмов (или новой группы) на основании анализа морфологических признаков. Однако сам подход к систематике остался при этом прежним.

Только в середине XX века Хеннигом были сформулированы принципиально новые идеи о том, что в основе классификации форм жизни должны лежать родственные, эволюционные отношения между ними: классификация должна быть филогенетической [Hennig, 1950]. Для этого подхода Хенниг предложил термин «кладистика» (cladistics), в отличие от традиционного, фенетического, подхода (phenetics). Так, копыта у парно- и непарнокопытных животных имеют различное эволюционное происхождение, соответственно — наличие копыт не дает никаких оснований для выделения группы копытных животных. Эволюционно парнокопытные гораздо более родственны, скажем, китам (отряд *Cetacea*), чем лошадям. Более подробно различия между кладистикой и фенетикой обсуждены в разд. 4.7.

Несмотря на то что кладистические идеи достаточно быстро получили широкое признание (хотя по целому ряду вопросов споры между сторонниками фенетического и кладистического подходов продолжают до сих пор), их практическое применение в биологии до начала молекулярной эры было ограничено. Понимание того, что генетические последовательности могут быть использованы для изучения филогении форм жизни, появилось сразу же после открытия структуры ДНК [Crick, 1958]. Однако только развитие методов быстрого получения и анализа большого объема генетической информации за последние два десятилетия предоставило широкие возможности для эволюционного анализа генетического материала. Подчеркнем, что мор-

фологические признаки также могут быть проанализированы с помощью кладистического подхода: различие между фенетикой и кладистикой определяется не типом анализируемых данных (морфологические, генетические и др.), но подходом к их анализу.

1.16. Проведение эволюционного анализа

В заключение этой главы кратко просуммируем ее ключевые положения и опишем логику и порядок проведения эволюционного анализа.

Генетическая информация в живых системах представлена молекулами нуклеиновых кислот, последовательностями нуклеотидов. При репликации этих молекул могут происходить ошибки, мутации, в том числе — замены, вставки и делеции нуклеотидов. В результате дочерние геномы будут отличаться от родительских. Разные мутации происходят с разной частотой, вероятностью. Эти мутации могут приводить к изменению приспособленности живой системы и подвергаться действию факторов естественного отбора, или быть эволюционно нейтральными. Кроме факторов естественного отбора, случайные, стохастические процессы (в значительной степени) определяют судьбу мутаций — их эволюционное закрепление или потерю.

Каждая родительская последовательность может давать несколько дочерних последовательностей, и мутации в этих дочерних последовательностях, вообще говоря, будут различными. Дальнейшая независимая эволюция этих дочерних последовательностей лежит в основе видообразования (дивергенция между эволюционными линиями).

Все эти эволюционные события происходят во времени. Чем больше времени прошло с момента разделения независимых эволюционных линий, тем, в общем случае, больше генетических различий будет наблюдаться между ними (концепция молекулярных часов).

В результате всех этих событий имеющаяся у живых систем генетическая информация будет различной, их нуклеотидные последовательности будут отличаться друг от друга.

Реконструкцией родственных отношений между формами жизни на основании анализа их генетических последовательностей, установлением законов эволюционных изменений наследственной информации занимается молекулярная филогенетика — ключевой раздел молекулярной эволюции. При филогене-

тическом анализе, гомологичные (имеющие общее эволюционное происхождение) генетические последовательности анализируемых форм жизни сравнивают между собой с помощью различных методов.

Поскольку при эволюции последовательностей происходят как замены, так и вставки-делеции нуклеотидов, то первым этапом эволюционного анализа в общем случае является выявление этих вставок-делеций. Этот этап анализа, называемый выравниванием последовательностей (гл. 2), необходим для того, чтобы установить гомологичные позиции анализируемых последовательностей.

После процедуры выравнивания последовательностей можно оценить число и качество различий между ними, рассчитать эволюционные дистанции (гл. 3). Анализируя генетические различия между последовательностями с помощью различных методов, можно реконструировать их эволюционную историю, родственные отношения между ними (собственно филогенетический анализ, гл. 4).

При проведении филогенетического анализа возникает ряд специальных вопросов, описываемых в гл. 5.

Эволюционный анализ генетической информации проводится при помощи многочисленных компьютерных программ с использованием различных баз генетических данных, описываемых в гл. 6.

В качестве самостоятельного раздела науки, со своими собственными задачами и методами исследований молекулярная филогенетика является ключевой составляющей более общего направления науки — биоинформатики, включающей в себя в том числе и сравнительную геномику, науку о взаимоотношениях между структурами и функциями макромолекул. Филогенетика и сравнительная геномика разделяют один и тот же объект исследования, макромолекулы во многом взаимосвязаны и опираются на результаты друг друга. В конце нашей книги приведен список книг для дополнительного чтения, включая и литературу по сравнительной геномике.