
بسم الله الرحمن الرحيم

وزارت علوم، تحقیقات و فناوری
دانشگاه تربت حیدریه

ژنومیک جمعیت در R
Population Genomics With R

امانوئل پارادیس

مسعود علی‌پناه

(دانشیار، گروه تولیدات گیاهی، دانشگاه تربت حیدریه)

(۱۴۰۴)

سرشناسه Paradis, Emmanuel
عنوان و نام پدیدآور : ژنومیک جمعیت در R/نویسنده امانوئل پارادیس؛ مترجم مسعود علی‌پناه.
مشخصات نشر : تربت حیدریه: دانشگاه تربت حیدریه، انتشارات، سال ۱۴۰۴
مشخصات ظاهری : ۴۲۱ ص
شابک : ۹۷۸-۶۰۰-۸۳۳۵-۴۱-۲
وضعیت فهرست نویسی : فیپا
یادداشت : عنوان اصلی: Population Genomics With R
یادداشت کتابنامه.
موضوع : آر (برنامه‌نویسی کامپیوتر)
R (Computer program language)
جمعیت
Population
فرازنگان شناسی
Metagenomics
شناسه افزوده : علی‌پناه، مسعود، ۱۳۴۸-، مترجم
شناسه افزوده دانشگاه تربت حیدریه
رده بندی کنگره : QA۲۷٦/٤٥
رده بندی دیویی : ۵۱۹/۵۰۲۸۵۵۱۳۳
شماره کتابشناسی ملی : ۱۰۱۱۶۸۸۶

این اثر مشمول قانون حمایت از مؤلفان و مصنفان و هنرمندان است. هرکس تمام یا قسمتی از این اثر را بدون اجازه ناشر، نشر، پخش یا عرضه کند مورد پیگرد قانونی قرار خواهد گرفت.

ژنومیک جمعیت در R
نویسنده: امانوئل پارادیس
مترجم: مسعود علی‌پناه.
چاپ اول
بها: تومان
نشانی ناشر: تربت حیدریه، کیلومتر هفت جاده مشهد، دانشگاه دولتی تربت حیدریه

مسئولیت کلیه مطالب این کتاب به عهده نگارنده می‌باشد. دانشگاه تربت حیدریه هیچگونه مسئولیتی در قبال صحت و سقم مطالب ندارد.

فهرست مطالب

فصل چهارم

فصل پنجم

فصل ششم

فصل هفتم

فصل هشتم

فصل نهم

فصل دهم

پیوست

پیشگفتار

برای سال‌های متمادی، ژنتیک جمعیت یک نظریه فوق‌العاده غنی و قدرتمند بود که عملاً هیچ حقایق مناسبی برای عمل‌کردن بر اساس آن وجود نداشت [...] ناگهان وضعیت تغییر کرد [...] و حقایق فراوان در قیف این ماشین تئوری ریخته شد. [...] کل رابطه بین نظریه و واقعیت‌ها نیاز به بازنگری دارد.

-لوونتین [۱۵۹]

گاهی ممکن است برخی از محققین بگویند که به‌صورت اتفاقی و پیش از صرف وقت و انرژی خود روی یک موضوع یا موضوعی شروع به کار کرده‌اند. این را می‌توان در مورد مشارکت من در ژنومیک جمعیت گفت. آموزش اولیه من در ژنتیک جمعیت بسیار سطحی بود و سابقه من در ژنومیک (یک رشته بسیار جدید در دوران دانشجویی) حتی سطحی‌تر بود. علاقه دیرینه من به تکامل و سرمایه‌گذاری من در R در اواخر دهه ۱۹۹۰ شروع شد، در نهایت باعث شد که بیشتر و بیشتر علاقه علمی خود را به ژنومیک جمعیت متمرکز کنم. با پشتوانه تجربه‌ام با ape، توسعه صفحات مجازی را شروع کردم که اولین‌بار در می ۲۰۰۹ منتشر شد. توسعه یک بسته R برای تجزیه‌وتحلیل تکاملی در این زمان چیز جدیدی نبود و چندین همکار، پروژه‌های مشابهی را آغاز کرده بودند. ایده این کتاب در برخی مباحثات با برخی از این همکاران بروز کرد. در این زمان توالی‌یابی با توان بالا (HTS) تازه شروع به پیشرفت خود کرده بود و ما تازه شروع به پیش‌بینی تأثیرات نهایی این انقلاب تکنولوژیکی کرده بودیم.

هکاتونی[۱] که در مرکز ملی سنتز تکاملی (NESCent) در دورهام، ایالات متحده، در مارس ۲۰۱۵ برگزار شد، برای من یک فرصت عالی برای توسعه ابزارهای جدید برای مدیریت داده‌های HTS

[۱] توضیح مترجم هکاتون یک رویداد کدنویسی اجتماعی است که برنامه‌نویسان برای یافتن راه‌حل‌های جدید و تکمیل پروژه‌های بزرگ مشترک، دور هم جمع می‌شوند.

در پگاس بود. پس از شروع گفتگو با جان کیمل در سپتامبر ۲۰۱۷، ایده Population Genomics With R به تدریج رشد کرد و به یک پروژه کتاب تبدیل شد.

سه ایده اصلی در پشت این کتاب وجود دارد. اولین مورد درنظرگرفتن انواع داده‌های ژنتیکی و ژنومی جمعیت، از ساده‌ترین داده‌های ژنتیکی تا بزرگ‌ترین داده‌های ژنومی در مقیاس بزرگ است. ایده دوم ارائه یک محیط محاسباتی واحد و مشترک برای پاسخگویی به طیف وسیعی از سؤالات یا رسیدگی به طیف گسترده‌ای از تجزیه‌وتحلیل با داده‌های ژنتیکی و ژنومی جمعیت است. ایده سوم ترویج استفاده از نرم افزارهای رایگان و متن باز است. در ادامه نوشتن، متوجه شدم که آماردانان و توسعه دهندگان ژنومیک از R بیشتر از آنچه فکر می‌کردم استفاده می‌کنند. پس از دفاع از استفاده از R برای بیش از دو دهه، این واقعیتی است که مطمئناً باعث رضایت من می‌شود.

مواد اولیه ژنومیک جمعیت با R، بسته‌های R هستند که در فصل ۱ فهرست شده‌اند. واضح است که این فهرست کاملی از منابع محاسباتی برای ژنومیک جمعیت نیست. من تا آنجا که ممکن است سعی کردم بسته‌هایی را در نظر بگیرم که عملیاتی باشند و در چارچوب کلی ژنومیک جمعیت که در بالا ذکر شد یکپارچه شوند. به علاوه، از ذکر بسته‌هایی که به‌وضوح نگهداری نمی‌شوند (به‌عنوان‌مثال، بسته‌های یتیم در CRAN) یا به نظر می‌رسد که به‌درستی کار نمی‌کنند، اجتناب کردم. در طول تحقیقاتم، مطمئناً بسته‌هایی را که باید در این کتاب گنجانده می‌شد، از دست داده‌ام. به‌عنوان‌مثال، DECIPHER، بسته‌ای که در BioConductor برای مدیریت پایگاه داده‌های بسیار بزرگ داده‌های توالی توزیع شده است، باید در چندین فصل از این کتاب ذکر می‌شد. از سوی دیگر، من بسته‌هایی را که روی سرور نیستند یا بیش از حد تخصصی هستند در نظر نمی‌گیرم: این بسته‌ها شامل چندین بسته R است که برای تجزیه‌وتحلیل جمعیت‌های انسانی توسعه یافته‌اند که تنها در صورت درخواست برای نویسندگان آن‌ها در دسترس هستند – اگرچه مشخص نیست این کار چگونه است. روش توزیع نرم‌افزار با چارچوب نرم‌افزار متن‌باز و آزاد که من سعی کردم از آن پیروی کنم در تضاد است.»

«نوشتن این کتاب یک فرآیند بسیار پیشرو بود و من از نظرات انتقادی و بسیار مفیدی درباره پیش‌نویس‌های اولیه اولیویه فرانسوا، سانتوش جیریراجان، سارا هندریکس و چندین منتقد ناشناس بهره بردم. هیلمار لاپ من را به هکاتونی که در سال ۲۰۱۵ در NESCent برگزار شد دعوت کرد،

جایی که یکی از هیجان انگیزترین هفته‌های کار و پیشرفت را سپری کردم: از او و همکاران و دوستانی که در آنجا بودند تشکر می‌کنم. تیبو جومبارت در بسیاری از موارد بحث‌های بسیار خوبی را به اشتراک گذاشت: از او برای سازماندهی و دعوت من به هکاتون دیگری در لندن تشکر می‌کنم. من این فرصت را داشتم که چندین کارگاه در مورد بسته‌هایی که توسعه می‌دهم برگزار کنم: این رویدادها تجربیات بسیار مفیدی هستند و من می‌خواهم از فردریک چیرولو، سولداد د استبان – تریویگنو، ژروم گودت و نیکلاس سالامین به‌خاطر دعوت‌هایشان تشکر کنم. از اگنس میگنو به‌خاطر حمایت کاملش از توسعه تحقیقات من در زمانی که رئیس مؤسسه علوم تکاملی در مونپلیه بود بسیار سپاسگزارم. من از جان کیمل بسیار سپاسگزارم که به من فرصت داد تا کتاب دیگری در مورداستفاده از R برای تجزیه‌وتحلیل تکامل بنویسم. رابین لوید استارکس با اشتیاق تمام جنبه‌های عملی دستنوشته من را بررسی کرد. در نهایت، من از همسرم سینتا و دخترم لور برای حمایت همیشگی‌شان سپاسگزارم.

بنگسان

امانوئل پارادیس

مارس ۲۰۲۰

توضیح نماد

E	موردانتظار
H	هتروزیگوسیتی
H	تنوع هاپلوئیدی
K	تعداد جمعیت، گروه یا خوشه
K	تعداد آلل یا جایگاه
L	احتمال
N	اندازه جمعیت
Ne	اندازه مؤثر جمعیت که با تعداد آلل‌های انتقال یافته به نسل بعد تعریف می‌شود
N	اندازه نمونه (تعداد افراد یا آلل)، تعداد ردیف در یک جدول یا در یک ماتریس
p	تعداد متغیرها، تعداد ستون‌ها در یک جدول یا در یک ماتریس
p_i	نسبت آلل i در جمعیت (i=۱,…,k)
$\hat{p}_i$	تخمین p_i
$\bar{p}_i$	مقدار پیش‌بینی p_i
Pr	احتمال
μ	میانگین، نرخ جهش
π	تنوع نکلئوتیدی
σ	انحراف معیار
σ^2	واریانس
$\sum$	ماتریس واریانس-کوواریانس
Θ	پارامتر ژنتیکی جمعیت تعیین می‌شود به‌عنوان

فصل اول

۱- مقدمه

۱-۱ وراثت، ژنتیک و ژنومیک

یکی از بزرگ‌ترین دستاوردهای زیست‌شناسی در طول قرن بیستم کشف مکانیسم‌های وراثت بود. به سختی می‌توان تمام نظریه‌هایی را که در طول قرن‌ها قبل از این کشف تدوین شده‌اند تصور کرد. امروزه، مارپیچ دوگانه ساختار DNA نماد علم است و DNA در حال حاضر طیف وسیعی از کاربردهای تکنولوژیکی و تجاری را دارد.

وراثت و مفاهیم مرتبط با آن عمیقاً در تاریخ بشریت ریشه دارد. ظهور کشاورزی در نقاط مختلف جهان بین ۱۰۰۰۰ تا ۵۰۰۰ سال پیش به‌وضوح با دانش در مورد وراثت برخی از گیاهان و حیوانات مرتبط شده است. طی هزاران سال، پرورش دهندگان عواقب وراثت را بر روی اشکال اهلی این گونه‌ها مشاهده کرده‌اند. در قرن نوزدهم، تحقیقات علمی وراثت با تعمیم مشاهدات میکروسکوپی، فرمول بندی قوانین وراثت توسط مندل، و کشف «نوکلئین» توسط میشر که بعدها به اسیدهای نوکلئیک تغییر نام داد، چرخش قابل‌توجهی پیدا کرد.

یکی از ویژگی‌های تاریخ ژنتیک که اغلب نادیده گرفته می‌شود این است که تقریباً هشت دهه طول کشید تا نشان داده شود که «DNA مسبب وراثت است و حتی آزمایش‌های درخشان اوری[۲] و همکارانش برای برخی از ژنتیک‌دانان که فکر می‌کردند وراثت با پروتئین‌ها رمزگذاری شده است، قانع‌کننده نبود. [۵۲]؛ بنابراین، ژنتیک جمعیت خیلی قبل از کشف حمایت فیزیکی از وراثت سرچشمه گرفت.

نشانه‌های تاریخی: وراثت، ژنتیک و ژنومیک

۱۸۶۶ :مندل قوانین وراثت خود را منتشر کرد [۱۸۴].

---

[۲] Avery

۱۸۶۹ :میشر DNA را کشف کرد [۴۷].
۱۹۴۴ . آوری و همکارن نشان داد که DNA پشتیبان وراثت است [۱۰].
۱۹۵۳ :واتسون و همکاران. ساختار مارپیچ دوگانه DNA را کشف کنید [۲۹۰].
۱۹۶۱ :کریک و همکاران. رمزگشایی کد ژنتیکی [۴۴].
۱۹۷۳ :گیلبرت و ماکسام اولین داده‌های توالی‌یابی DNA را منتشر کردند [۹۵].
۱۹۸۴ :کشف ریزماهواره‌ها [۲۹۵].
۱۹۹۶ :اولین فناوری توالی‌یابی با توان بالا [۲۳۷].
۲۰۰۱ :اولین ژنوم انسان منتشر شد [۱۲۷].
۲۰۱۰: تکمیل فاز اول پروژه ۱۰۰۰ ژنوم [۲۷۰].

در طول قرن بیستم، روش‌هایی که زیست‌شناسان برای مطالعه وراثت و بعداً DNA استفاده کردند، به تدریج قدرت خود را افزایش دادند (به فصل ۲ مراجعه کنید). رشد فن‌آوری‌های توالی‌یابی با توان بالا عامل بسیار مهمی در توسعه ژنومیک جمعیت بوده است. ژنومیک در دهه اخیر به‌عنوان یک رشته علمی و موضوعی با «علاقه اجتماعی قابل توجه» اهمیت قابل‌توجهی یافته است. این پیشرفت بر حوزه ژنتیک جمعیت نیز تأثیر گذاشته است.

این کتاب تعاریف زیر را اتخاذ می‌کند. ژنتیک جمعیت مطالعه تنوع در ژنوتیپ‌ها در بین افراد در مکان و زمان، از جمله نیروهای پشت این تنوع است. ژنومیکس مطالعه ساختار و عملکرد ژنوم است. ژنومیک جمعیت مشابه ژنتیک جمعیت، اما برای تعداد بسیار زیادی از جایگاه‌ها است، معمولاً در کل ژنوم یک گونه اعمال می‌شود؛ بنابراین، ژنومیک جمعیت را می‌توان به‌عنوان یک نسخه "مقیاس شده" از ژنتیک جمعیت در نظر گرفت که با حداقل تعداد زیادی مکان تا کل ژنوم گونه مورد نظر سروکار دارد [۲۰].

نشانه‌های تاریخی: ژنتیک جمعیت

۱۹۳۰ :انتشار نظریه ژنتیکی فیشر در مورد انتخاب طبیعی [۷۷].
۱۹۴۹ :انتشار مقاله رایت در مورد ساختار ژنتیکی جمعیت [۳۰۳].
۱۹۵۵ :مقاله کیمورا در مورد تثبیت آلل تحت رانش ژنتیکی [۱۴۲].

۱۹۶۶ :مطالعات تجربی اهمیت تنوع مولکولی را در جمعیت‌های طبیعی نشان داد [۱۰۷، ۱۶۰].

۱۹۸۲ :کینگمن سه مقاله تأسیسی در مورد کوالسنت منتشر کرد [۱۴۷].

۲۰۰۵: انتشار سریالی زنجیره مارکوف که تجزیه‌وتحلیل داده‌های ژنومی را با نوترکیب تسهیل کرد [۱۸۲].

۱-۲ اصول ژنومیک جمعیت

این بخش با توضیحاتی در مورد واحدهای مورداستفاده در این کتاب آغاز می‌شود. معانی بیولوژیکی برخی از اصطلاحات استفاده شده در اینجا (بازها، دو رشته‌ای، و ...) در ادامه در زیر بخش توضیح داده شده است.

۱-۲-۱ واحدها

واحد اصلی ژنوم باز است، بخشی از نوکلئوتید که متغیر است: نماد آن "b" است. ژنوم‌ها می‌توانند کوچک یا (بسیار) بزرگ باشند، بنابراین استفاده از پیشوندهای وام گرفته شده از سیستم بین‌المللی واحدها برای بیان اندازه یک ژنوم یا طول یک توالی DNA رایج است:

یک کیلوباز= ۱kb= ($۱۰^{۳}$) ۱۰۰۰ باز

یک مگا باز= ۱Mb= ($۱۰^{۶}$) ۱۰۰۰۰۰۰ باز

یک گیگا باز= ۱Gb= ($۱۰^{۹}$) ۱۰۰۰۰۰۰۰۰۰ باز

یک ترا باز= ۱Tb= ($۱۰^{۱۲}$) ۱۰۰۰۰۰۰۰۰۰۰۰ باز

یک پتا باز= ۱Pb= ($۱۰^{۱۵}$) ۱۰۰۰۰۰۰۰۰۰۰۰۰۰ باز

توجه داشته باشید که "باز (ها) " اغلب به معنای واقعی " جفت باز (ها) " استفاده می‌شود. زیرا DNA تقریباً همیشه دو رشته‌ای است. اگرچه این متناقض است، اما «bp» معمولاً به‌عنوان نماد به‌جای «b» در صورت پیشوند استفاده نمی‌شود، به‌عنوان‌مثال: ۱۰۰۰ kb ۱ = bp.

ژنومیکس مدرن به‌شدت با علم کامپیوتر مرتبط است، به‌طوری‌که ما اغلب نیاز داریم به کمیت اطلاعات، استفاده از حافظه یا اندازه فایل اشاره کنیم. واحد اصلی اطلاعات بیت

(متغیر باینری) و واحد عملی بایت با نماد "B" (یک بایت = هشت بیت) است. رایج‌ترین واحدهای استفاده از حافظه عبارتند از:

یک کیلوبایت= ۱kB= $۱۰^{۳}$ بایت

یک مگا بایت= ۱MB= $۱۰^{۶}$ بایت

یک گیگا بایت= ۱GB= $۱۰^{۹}$ بایت

یک ترا بایت= ۱TB= $۱۰^{۱۲}$ بایت

در این کتاب، ما از واحدهای کوچک جرم نیز استفاده خواهیم کرد، زیرا DNA معمولاً در مقادیر بسیار کم وجود دارد (به فصل ۲ مراجعه کنید):

یک میکروگرم = $۱\mu g$= $۱۰^{-۶}$ گرم

یک نانوگرم. = $۱\mu g$= $۱۰^{-۹}$ گرم

یک پیکوگرم = $۱\mu g$= $۱۰^{-۱۲}$ گرم

۱-۲-۲ ساختارهای ژنوم

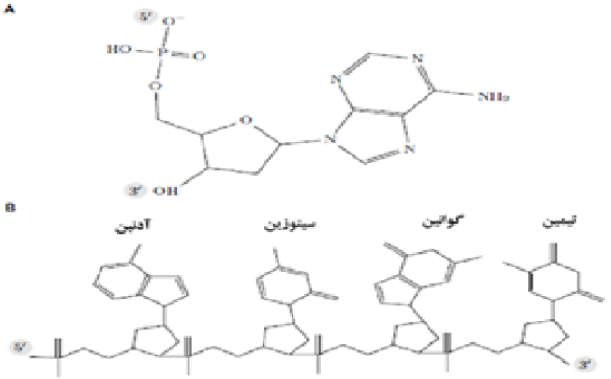
DNA پلیمری است که از تکرار نوکلئوتیدها ساخته شده است که خود از سه مولکول تشکیل شده‌اند: فسفات، دئوکسی ریبوز و یک باز (شکل ۱-۱). پایه یک نوکلئوتید می‌تواند آدنین (A)، سیتوزین (C)، گوانین (G) یا تیمین (T) باشد. نام "باز" از این واقعیت ناشی می‌شود که این مولکول‌ها در محلول بازی هستند (یعنی یون‌های هیدروکسید OH- را آزاد می‌کنند، برخلاف اسیدهایی که یون‌های هیدروژن H+ را آزاد می‌کنند). در واقع بازهای زیادی در طبیعت وجود دارد (به‌عنوان‌مثال، کافئین، گزانتین)، اما تنها این چهار باز در DNA یافت می‌شوند. توالی این بازها در یک پلیمر DNA، اطلاعات ژنتیکی موردنیاز برای انجام عملکردهای اساسی حیات، مانند کدگذاری توالی پروتئین‌ها یا کدگذاری توالی‌های تنظیم‌کننده را ذخیره می‌کند.

DNA تقریباً همیشه دو رشته‌ای است به‌گونه‌ای که پایه‌های هر دو پلیمر (رشته) جفت‌های خاصی را تشکیل می‌دهند A :با T و G با C (شکل ۱-۲). این دو رشته توسط نیروهای

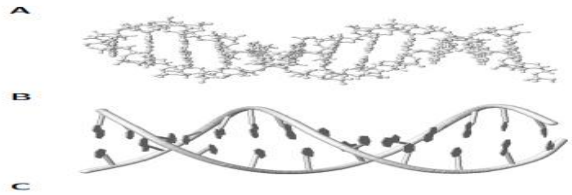
ضعیفی که الکترون‌ها را بین بازهای یک جفت به اشتراک می‌گذارند، محدود می‌شوند: دو الکترون برای یک جفتA-T ، سه الکترون برای یک جفت.G-C

چند استثنا در قاعده DNA به‌عنوان پشتیبان اطلاعات ژنتیکی وجود دارد: در برخی ویروس‌ها، اسید ریبونوکلئیک (RNA) پشتیبان اطلاعات است. RNA مشابه DNA است؛ اما با دو تفاوت: اوراسیل (U) به‌جای T و ریبوز به‌جای دئوکسی ریبوز استفاده می‌شود (شکل ۱-۳).

دزوکسیداسیون ریبوز (حذف یک اتم اکسیژن) باعث می‌شود DNA از نظر شیمیایی واکنش‌پذیری کمتری داشته باشد و در نتیجه بیشتر از RNA جدولی باشد. در واقع، ویروس‌ها می‌توانند دارای ژنوم‌های ساخته شده از DNA یا RNA، تک یا دو رشته‌ای باشند. در تمام اشکال زنده دیگر، ژنوم همیشه از DNA دو رشته‌ای ساخته شده است.



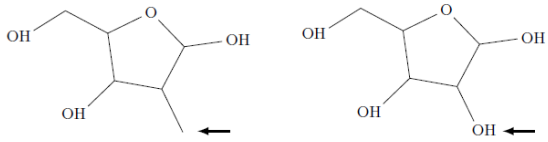
شکل ۱-۱ (A) یک نوکلئوتید ساخته شده از یک فسفات (HPO_4)، یک دئوکسی ریبوز ($C_5H_{10}O_4$) و یک باز (در اینجا آدنین، $C_5H_5N_5$). حاشیه‌نویسی ۵ و ۳ نشان می‌دهد که نوکلئوتیدها کجا به یکدیگر متصل می‌شوند تا DNA تک رشته‌ای بسازند. (ب) یک مولکول DNA تک رشته‌ای ساخته شده از چهار نوکلئوتیدها با توالی ACGT.



شکل ۲-۱ (A) یک مولکول DNA جفت دوازده باز که اتم‌های آن را نشان می‌دهد. (ب) نمایشی از همان مولکول که جفت‌های باز (چند ضلعی‌های خاکستری تیره) و ستون فقرات ساخته شده از فسفات و دزوکسی ریبوز (لوله‌های خاکستری روشن) را نشان می‌دهد. (ج) دوازده جفت باز این مولکول (A و B با http://jena۳d.leibniz-fli.de/ ترسیم شده‌اند).

RNAها در واقع مولکول‌های بسیار مهمی در موجودات زنده هستند. یک مرحله میانی برای بیان اطلاعات ذخیره شده در DNA ، سنتز RNA یا رونویسی است (شکل ۱.۴). برخی از مولکول‌های RNA در سنتز پروتئین استفاده می‌شوند و برخی دیگر نقش‌های متفاوتی در سلول دارند (شکل ۱-۵).

ژنوم‌ها می‌توانند اندازه‌ها و ساختارهای بسیار متفاوتی داشته باشند (جدول ۱-۱). ویروس‌ها در واقع ساده‌ترین ژنوم‌ها را دارند، معمولاً ۲ تا ۵۰ کیلوبایت طول دارند، اما برخی می‌توانند به ۲.۵ مگابایت برسند [۲۲۲]. پروکاریوت‌ها موجودات تک سلولی با ژنوم نسبتاً ساده هستند: کوچک‌ترین آنها کمی بزرگ‌تر از ۱۰۰ کیلوبایت و بزرگ‌ترین آنها بیش از ۱۲ مگابایت هستند. به استثنای چند مورد، ویروس‌ها و پروکاریوت‌ها این ویژگی را دارند که ژنوم آنها از یک مولکول DNA (یا RNA برای برخی ویروس‌ها) ساخته شده است. پروکاریوت‌ها مولکول‌های DNA اضافی به نام پلاسمید دارند که در ژنوم اصلی‌شان ادغام نشده‌اند و به‌طور مستقل تکثیر می‌شوند.



شکل ۱-۳ دزوکسی ریبوز (سمت چپ) و ریبوز (راست) مولکول‌های بسیار مشابهی هستند: اکسیژن ازدست‌رفته در اولی باعث می‌شود که از نظر شیمیایی پایدارتر از دومی باشد.

یوکاریوت‌ها پیچیده‌تر از پروکاریوت‌ها هستند. یک موجود ممکن است از چندین سلول ساخته شده باشد، اگرچه بسیاری از گونه‌های یوکاریوت‌ها تک سلولی هستند (آغازیان). ژنوم آنها در چندین مولکول DNA مملو از پروتئین برای ساخت کروموزوم‌ها مرتب شده

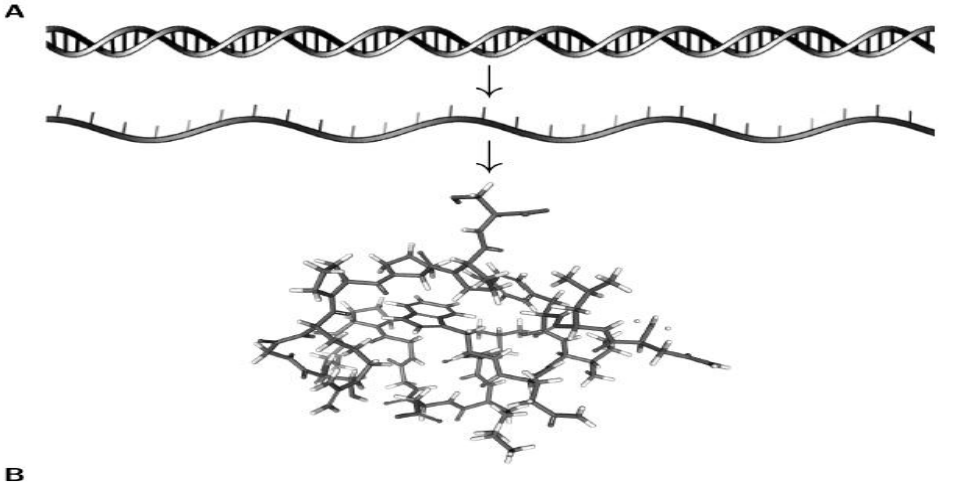
است. تعداد کروموزوم‌ها بسیار متفاوت است: معمولاً بین چند تا چند ده است Protist Oxytricha trifallax یک مورد قوی است. این ارگانیسم تک سلولی بزرگ دارای دو هسته است: ماکرونوکلئوس حاوی ژنوم جسمانی با ۵۰ مگابایت پخش شده بر روی ۱۵۶۰۰ کروموزوم و میکرونوکلئوس با ژنوم ۵۰۰ مگابایت تکه تکه شده به بیش از ۲۲۵۰۰۰ مولکول DNA مورداستفاده برای تولیدمثل جنسی [۳۷، ۲۶۱]. اندازه مولکول DNA یک کروموزوم یوکاریوتی بسیار متفاوت است: از چند ۱۰۰ جفت باز (نانوکروموزوم) تا چند ۱۰۰ مگابایت.

یوکاریوت‌ها، مانند پروکاریوت‌ها، ژنوم‌های جانبی دارند، اما به‌جای «آزاد» بودن در سلول، در اندامک‌های خاصی مانند میتوکندری‌های موجود در اکثر سلول‌های یوکاریوتی یا کلروپلاست‌های گیاهان فتوسنتزی قرار دارند. آپیکوپلاست یک اندامک خاص برای برخی از آغازیان است که ژنوم کوچکی نیز دارد (جدول ۱.۱).

تفاوت‌های بسیار دیگری بین ژنوم‌های یوکاریوتی و پروکاریوتی وجود دارد: دو مورد از آنها در اینجا قابل ذکر است. اولاً، سازماندهی توالی‌های کدکننده (آنهایی که به mRNA رونویسی می‌شوند) در پروکاریوت‌ها ساده است، جایی که آنها برای یک پروتئین معین پیوسته هستند. از سوی دیگر، در یوکاریوت‌ها، بخش‌های کدکننده (اگزون‌ها) ناپیوسته هستند و با بخش‌های غیر کدکننده پراکنده هستند (اینترون‌ها، شکل ۱-۶). دوم، ژنوم‌های یوکاریوتی اغلب در چندین نسخه در یک سلول وجود دارند، معمولاً دو نسخه (دیپلوئیدی)، و بیشتر گونه‌ها با تناوب مراحل هاپلوئید و دیپلوئید مشخص می‌شوند که در آن کروموزوم‌های همولوگ بخش‌هایی از DNA خود را در طول انتقال از دیپلوئید به سلول مبادله می‌کنند. فاز هاپلوئید (به زیر مراجعه کنید).

بسیاری از گونه‌های یوکاریوتی بیش از دو نسخه از ژنوم خود در یک سلول دارند، پدیده‌ای به نام پلی پلوئیدی. این در واقع بسیار رایج‌تر از آنچه تصور می‌شود است و در برخی سلول‌های خاص بیشتر ارگانیسم‌های چند سلولی، چه به‌طور طبیعی یا پاتولوژیک، رخ می‌دهد [۲۴۹، ۲۸۱]. رایج‌ترین وضعیت موجودات پلی پلوئیدی تتراپلوئیدی (چهار نسخه

از کروموزوم‌ها) است، اما موقعیت‌های مختلفی را می‌توان مشاهده کرد. گونه‌هایی با تعداد فرد کپی نادر هستند و عموماً به‌صورت کلونی تکثیر می‌شوند. به‌عنوان‌مثال، درختچه سه گانه Lomatia tasmanica در کمتر از ۲ کیلومتر مربع زندگی می‌کند که به‌شدت در معرض خطر است [۱۷۵]. از سوی دیگر، خرچنگ مرمری، Procambarus virginalis، سه پلوئیدی و مهاجم در ماداگاسکار است، جایی که گونه‌های بومی خرچنگ را تهدید می‌کند [۱۰۲].



شکل ۱-۴ نحوه استفاده از اطلاعات ژنتیکی برای کدگذاری صفت فنوتیپی مانند پروتئین. (الف) DNA دو رشته‌ای به mRNA رونویسی می‌شود که سپس به پروتئین ترجمه می‌شود. (ب) همان فرآیند نشان‌داده‌شده به‌عنوان توالی از بازها و اسیدهای آمینه، در اینجا یکی از کوچک‌ترین پروتئین‌های شناخته شده [۱۲، ۲۰۰] (توالی اسیدآمینه و تصویر سه بعدی از https://www.rcsb.org/۳d-view /۱L۲Y با استفاده از NGL Viewer[۲۳۸]). "کد سه حرفی برای اسیدهای آمینه در اینجا استفاده شده است.

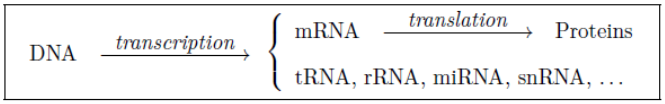
جدول ۱-۱ اندازه و ساختار معمول ژنومی

	اندازه معمول ژنوم	تعداد کپی
ویروس	۱۰kb	هاپلوئید
پروکاریوت	۱Mb	هاپلوئید
پلاسمیدها	۱kb	هاپلوئید
یوکاریوت‌ها	۱Gb	هاپلوئید، دیپلوئید، پلی پلوئید
میتوکندری	۱۶kb	هاپلوئید
اپیکوپلاست	۳۵kb	هاپلوئید
کلروپلاست	۱۰۰kb	هاپلوئید

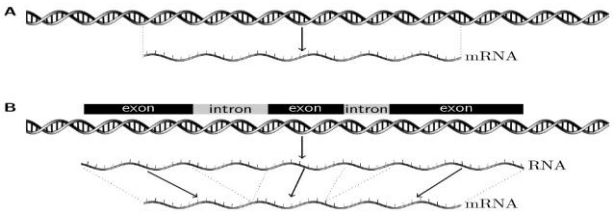
DNA از نسلی به نسل دیگر (تقریباً) دقیق کپی و منتقل می‌شود. برای انتقال اطلاعات ژنتیکی دو روش تولیدمثل وجود دارد: کلونی و جنسی. در تولیدمثل کلونال، یک والد تنها یک یا چند فرزند تولید می‌کند که یک کپی دقیق از ژنوم والدین دارند که می‌تواند هاپلوئید، دیپلوئید یا پلی پلوئید باشد. در تولیدمثل جنسی، والدین دیپلوئیدی گامت‌های هاپلوئید تولید می‌کنند: دو گامت (از یک فرد یا از دو فرد) به هم ملحق می‌شوند تا یک فرزند دیپلوئید تولید کنند. در طول تولید گامت‌ها، کروموزوم‌های همولوگ، DNA را برای تولید ترکیب‌های جدیدی از توالی‌ها مبادله می‌کنند: این نوترکیبی است. فرآیند نوترکیبی ژنتیکی نیز در پروکاریوت‌ها و ویروس‌ها وجود دارد اما به شکلی متفاوت. تولیدمثل جنسی را می‌توان در ارگانیسم‌های پلی پلوئید نیز مشاهده کرد: افراد تتراپلوئید معمولاً گامت‌های دیپلوئیدی تولید می‌کنند [۲۵۰].

تولیدمثل جنسی فقط در یوکاریوت‌ها یافت می‌شود درحالی‌که تولیدمثل کلونال در همه گروه‌ها مشاهده می‌شود - البته نه در همه گونه‌ها. در بسیاری از گروه‌های دارای تولیدمثل جنسی (مثلاً مهره داران یا بی مهرگان)، مرحله هاپلوئید رشد نمی‌کند (در حالت تک سلولی باقی می‌ماند) و افراد بیشتر عمر خود را به‌صورت دیپلوئید می‌گذرانند. یک استثنای قابل توجه از این قاعده توسط گیاهان متعلق به Pteridophyta (سرخس) است که دیپلوئید (اسپوروفیت) هستند و دارای مرحله هاپلوئید (گامتوفیت) هستند که برای تولید گامت رشد می‌کنند. Protist Plasmodium falciparum .بیشتر هاپلوئید است و به‌صورت کلون در

انسان تکثیر می‌شود (جایی که باعث مالاریا می‌شود) درحالی‌که تولیدمثل جنسی در پشه‌های جنس آنوفل رخ می‌دهد.



شکل ۱-۵ خلاصه‌ای از نحوه استفاده از اطلاعات ذخیره شده در DNA در موجودات زنده. mRNA: RNA پیام رسان، tRNA : RNA انتقالی، rRNA : RNA ریبوزومی، miRNA: میکرو RNA، snRNA RNA هسته‌ای کوچک.



شکل ۱-۶ (A) یک ژن به mRNA رونویسی می‌شود که بعداً به پروتئین (در ویروس‌ها، پروکاریوت‌ها و اندامک‌های یوکاریوتی) ترجمه می‌شود. (ب) بیشتر ژن‌های یوکاریوتی از مناطق کدکننده (اگزون) و غیر کدکننده (اینترون) ساخته شده‌اند و سنتز mRNA یک فرآیند دو مرحله‌ای است.

۱-۲-۳ جهش‌ها

جهش از مدت طولانی شناخته شده است: از آغاز کشاورزی، پرورش دهندگان قادر به تولید لاین‌هایی از گیاهان یا حیوانات با ویژگی‌های پایدار و ثابت در طول نسل بوده‌اند. بااین‌حال، حتی در دقیق‌ترین شرایط، برخی از افراد با ویژگی (ها) غیرمنتظره به طور پراکنده در این لاین‌های پرورش ظاهر می‌شوند. اخیراً، زیست شناسان توانستند وقوع جهش‌هایی را در آزمایشگاه‌ها با باکتری‌ها یا مگس‌های میوه مشاهده کنند.

برای مدت طولانی، ژنتیک جمعیت یک جهش را تغییر از یک آلل به آلل دیگر بدون هیچ فرض دیگری در مورد ماهیت چنین تغییراتی در نظر می‌گرفت. با ظهور ژنتیک مولکولی، کلمه جهش به معنای تغییر در مولکول‌های DNA منتقل شده از طریق نسل‌ها مستقل از اثرات فنوتیپی بالقوه آن تغییر یافته است. به طور مشابه، واژه‌های جایگاه (محل قرارگرفتن یک ژن در کروموزوم) و آلل (انواع مختلف یک ژن) در ژنومیک جمعیت معانی کمی

متفاوت دارند: جایگاه بخشی از ژنوم است که چندشکلی را نشان می‌دهد، درحالی‌که آلل‌ها توالی‌های مختلفی مشاهده شده برای این جایگاه هستند.

جدول ۱-۲ فراوانی جهش‌های (درصد) مشاهد شده در بین ۲۵۰۴ ژنوم انسان و ۱۱۳۵ ژنوم شامپانزه تال.

نوع جهش	انسان	آرابیدوپسیس تالیانا
SNP دو آللی	۹۵.۵۳	۸۸.۲۶
درج – حذف(indels)	۴.۰۷	۱۱.۷۴
SNPهای چند آللی	۰.۳۳	-
انواع ساختاری	۰.۰۷	-

جهش‌ها در سطح DNA را می‌توان به پنج دسته اصلی طبقه‌بندی کرد:

- جهش‌های تک نوکلئوتیدی،
- درج – حذف،
- بازآرایی ژنوم،
- دوبلیکیشن ژن،
- دبلیکیشن ژنوم.

اولین دسته رایج‌ترین است: باعث تغییر یک باز به باز دیگر می‌شود. یکی از پیامدهای این جهش‌ها، در سطح جمعیت، وجود پلی مورفیسم تک نوکلئوتیدی یا SNP است. در عمل، اصطلاح SNP معمولاً به مواردی محدود می‌شود که فقط دو آلل در یک مکان خاص مشاهده شود. اگر سه یا چهار آلل مشاهده شود، می‌توان در مورد SNP یا MNP چند آللی صحبت کرد. برای جلوگیری از سردرگمی، باید از «SNP biallelic» یا «SNP» دقیق» استفاده کرد تا تأکید شود که فقط دو آلل مشاهده شده است.

درج‌ها - حذف‌ها (indels) منجر به افزایش یا از دست دادن نوکلئوتیدها، معمولاً تعداد کمی هستند. آنها دومین نوع جهش شایع هستند. دسته‌بندی‌های دیگر جهش‌ها در مجموع به‌عنوان انواع ساختاری شناخته می‌شوند: آنها منجر به تغییرات چشمگیرتر در ژنوم می‌شوند و بسیار کمتر هستند.

نوشته‌های منتشرشده در سال ۲۰۱۵ از توالی‌های ۲۵۰۴ ژنوم انسانی از ۲۶ جمعیت در آفریقا، آسیا، اروپا و آمریکا [۲۷۲] معیاری برای ارزیابی فراوانی انواع مختلف جهش‌ها ارائه کرد.

از مجموع ۳,۲۴۱,۹۵۳,۴۲۹ باز (طول ژنوم مرجع GRCh۳۸ مورداستفاده در این مطالعه)، ۸۸,۳۳۲,۰۱۵ گونه ژنتیکی (یا جایگاه) شناسایی شد. SNP ها و ایندل‌های دقیق بیش از ۹۹.۵ درصد از پلی مورفیسم مشاهده شده را نشان می‌دهند (جدول ۱-۲). این واقعیت که MNPها تقریباً ۳۰۰ برابر کمتر از فقط SNP ها بودند، ممکن است نتیجه کوچک بودن جمعیت مؤثر جمعیت انسانی باشد [۱۳۶] (به بخش بعدی مراجعه کنید). یکی دیگر از پیامدهای این امر، شباهت ژنومی بسیار قوی انسان‌ها است: دو فرد به طور تصادفی انتخاب شده، ۹۹.۹ درصد از ژنوم‌های مربوطه خود را یکسان دارند [۲۷۲].

یک مطالعه در مقیاس بزرگ مشابه بر اساس ۱۱۳۵ ژنوم شاهی تال (Arabidopsis thaliana)، ۱۳۵، ۹۷۵نوع از ۱۱۹،۶۶۷،۷۵۰ پایه را نشان داد. با یک مکان متغیر در هر ۱۰ جفت باز، این متراکم‌ترین ژنوم یوکاریوتی است که تاکنون از نظر گونه‌های طبیعی شناخته شده است [۲۷۳]. فقط SNP های دو آللی و ایندل‌های کوچک (≤ ۴۰ جفت باز) در این مطالعه ارزیابی شدند (جدول ۱-۲). یک آزمایش اصلاحی روی این گیاه همراه با تعیین توالی ژنوم نشان داد که تعویض‌های تک پایه شایع‌ترین جهش در ژنوم است [۲۰۶]. با این وجود، فناوری‌های مختلف، نمونه‌برداری‌ها و ویژگی‌های زیست‌شناختی مورداستفاده در این مطالعات، تعمیم‌ها را دشوار می‌کند و بسیاری از سؤالات هنوز باز هستند - اگرچه به برخی از آنها پاسخ داده می‌شود. مطالعات اخیر بر روی جمعیت‌های انسانی نشان داد که در مقایسه با آنچه چند سال پیش نشان داده شد، تعداد SNP ها در ژنوم انسان به مراتب بیشتر است (چند صد میلیون که بیشتر آنها دارای یک آلل بسیار نادر هستند) [۲۶۴، ۲۶۸]، و که انواع ساختاری گسترده‌تر هستند [۳۱، ۲۵۱]. مطالعه تنوع ژنومی در جمعیت‌های طبیعی هنوز در حال انجام است و مطمئناً حقایق شگفت‌انگیز جدیدی را پس از انتشار داده‌های سایر افراد و جمعیت‌ها آشکار خواهد کرد.

۱-۲-۴ رانش و انتخاب

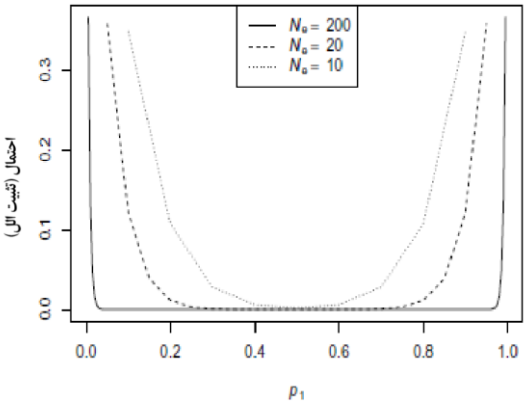
رانش ژنتیکی فرآیندی است که در آن فراوانی آللی با نمونه‌برداری تصادفی آلل‌ها از نسلی به نسل دیگر تغییر می‌کند. رانش همیشه در جمعیت‌های طبیعی وجود دارد، حتی اگر آنها

بزرگ یا در حال گسترش باشند، اگرچه در جمعیت‌های کوچک قوی‌تر است. یک راه ساده برای نگاه کردن به رانش، درنظرگرفتن یک مکان منفرد با دو آلل در جمعیتی با اندازه جمعیت مؤثر Ne است که در آن بسامدهای آلل p۱ و p۱-۱=p۲ است. می‌خواهیم به این سؤال پاسخ دهیم: اگر اندازه جمعیت ثابت باشد، احتمال اینکه یک آلل در نسل بعدی از بین برود چقدر است؟ اگر فرض کنیم که انتخاب تصادفی است (انتخابی وجود ندارد)، آنگاه پاسخ را می‌توان با استفاده از احتمالات دوجمله ای نمونه‌برداری از آلل‌ها با پارامترهای Ne و p۱ پیدا کرد (شکل ۱-۷). واضح است که احتمال تثبیت آلل در صورت متعادل بودن فراوانی آلل بسیار کم است (p۱=p۲=۰,۵) حتی اگر Ne بسیار کوچک باشد. بااین‌حال، در این حالت کاملاً بعید است که این فراوانی‌ها پایدار باشند (در واقع احتمال p۱=p۲=۰,۵ در نسل بعدی ۰.۲۵ است)؛ بنابراین، به ناچار p۱ به سمت چپ (راست) محور xشکل ۱-۷ می‌رود و در نتیجه، احتمال تثبیت آلل در طول زمان افزایش می‌یابد.

کیمورا و اوتا [۱۴۶] نشان دادند که زمان موردانتظار برای تثبیت یک آلل خنثی جدید در یک جمعیت متناسب با ۴□□ است. یکی دیگر از نتایج اساسی در مورد رانش را کیمورا و کرو [۱۴۵] ارائه کردند که آن‌ها نشان دادند تعداد آلل‌های موردانتظار در یک جمعیت ۱ + ۴μNe است جایی که μ، نرخ جهش است؛ بنابراین، یک جمعیت بزرگ‌تر می‌تواند حاوی آلل‌های بیشتری باشد، اما این به میزان جهش نیز بستگی دارد.

انتخاب یکی دیگر از مکانیسم‌های اصلی است که منجر به تغییر در فراوانی آلل در جمعیت‌های طبیعی می‌شود. به طور کلاسیک، سه نوع اساسی انتخاب در نظر گرفته می‌شود (شکل ۱.۸) [۱۵۶، ۲۳۱]. انتخاب مثبت منجر به افزایش فراوانی آلل‌های انتخابی سودمند می‌شود، حتی اگر در جمعیت دارای فراوانی پایین باشند. انتخاب برای خالص‌سازی منجر به کاهش یا حذف آلل‌های نامطلوب انتخابی می‌شود که فراوانی کمی در جمعیت دارند. انتخاب برای متعادل کردن (یا متنوع کردن) منجر به حفظ چندین آلل در جمعیت می‌شود که ممکن است مزایای انتخابی در موقعیت‌ها یا مکان‌های مختلف داشته باشند. سایر اشکال انتخاب سطوح انتخاب را در نظر می‌گیرند [۲۰۴]. مدت‌ها قبل از اینکه DNA به‌عنوان عامل

وراثت شناخته شود، انتخاب طبیعی یک انگیزه اولیه در ژنتیک نظری جمعیت بوده است. فیشر در یکی از فصل‌های خود نتیجه‌گیری کرد که «شالوده تئوری بقا [تغییر تکاملی] انتخاب طبیعی است» [۷۷]. با توالی‌های DNA، انتخاب را می‌توان در سطح مولکولی ارزیابی کرد (شکل ۱-۹). فصل ۱۰ روش‌های تجزیه‌وتحلیل توالی‌های DNA یا داده‌های ژنومی را بررسی می‌کند.

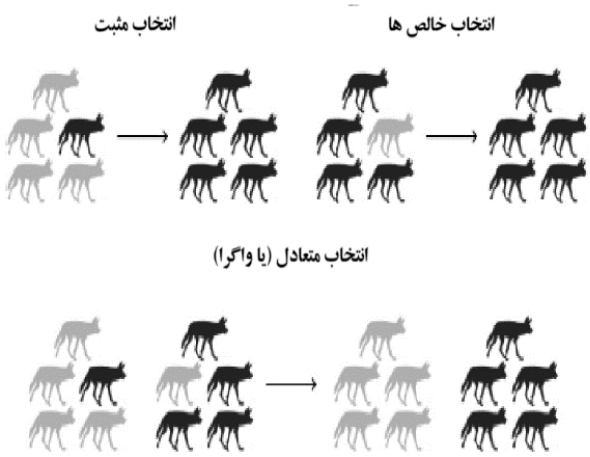


شکل ۱-۷ احتمال از دست دادن یک آلل (یعنی تثبیت آلل) در یک مکان دو آللی در یک نسل باتوجه‌به فراوانی یکی از دو آلل (p_1) و اندازه جمعیت مؤثر (Ne).

۱-۳ بسته‌ها و کنوانسیون‌های R

جدول ۱-۳ فهرست بسته‌های اصلی R را که در این کتاب استفاده می‌شوند و جدول ۱-۴ فهرست بسته‌هایی که بیشتر در یک فصل استفاده می‌شوند را نشان می‌دهد. اکثر این بسته‌ها در شبکه جامع آرشیوی R (CRAN) که منبع اصلی بسته‌های R است، توزیع می‌شوند. BioConductor مخزن دیگری از بسته‌های تخصصی R در بیوانفورماتیک است. این بسته‌ها فهرست کاملی از منابع R برای ژنومیک جمعیت نیستند، بلکه مجموعه‌ای از بسته‌هایی هستند که در مجموع ادغام می‌شوند تا یک «اکوسیستم» نرم‌افزاری برای ژنومیک جمعیت ادغام شده در R [۲۱۵] ایجاد کنند. علاوه بر اینها، از چند بسته اضافی برای بررسی داده‌های جغرافیایی استفاده می‌شود (ص ۲۴۲). alleHap، GeneImp (هر دو در ص ۱۶۸)،

pophelper (ص. ۲۱۸)، aphid (ص. ۲۹۲)، GENESIS (ص. ۳۱۴) و MINOTAUR (ص. ۳۲۳) نیز به اختصار مورد بحث قرار گرفته‌اند. پیوست A نحوه نصب این بسته‌ها را توضیح می‌دهد.



شکل ۱-۸ انواع اصلی انتخاب طبیعی در جمعیت‌ها

در این کتاب، نام توابع و سایر اشیاء در R با فونت monospace چاپ شده است. از پرانتزها برای تشخیص توابع از اشیاء دیگر استفاده می‌شود، مگر اینکه ابهامی وجود داشته باشد، به‌عنوان‌مثال: " ()print" برای "چاپ تابع." استفاده می‌شود. نام بسته‌ها با فونت sans serif چاپ می‌شود. دستورات R با اعلان معمول "بیشتر از" نشان داده می‌شوند درحالی‌که در دستورات سیستم با دستور "دلار" نشان داده می‌شوند:

```
> ls() # this is an R command
```

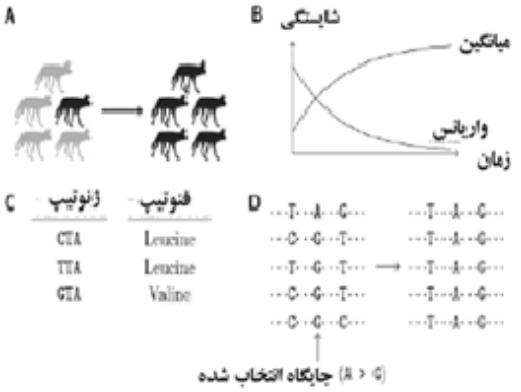
جدول ۱-۳ بسته‌های اصلی استفاده شده در این کتاب. به جز موارد ذکر شده، همه بسته‌ها بر روی CRAN هستند

نام	عنوان	منبع
Adegenet	ژنتیک جمعیت اسپاتیال و چندمتغیره	[۱۳۲]
ape	تجزیه فیلوژنتیک و تنوع	[۲۱۶]
Biostrings[a]	زیست استرینگ	[۲۰۸]
pegas	ژنومیکس جمعیت و تکاملی	[۲۱۰]
SNPRelate[a]	دسته ابزار محاسبات موازی برای داده‌های SNP	[۳۱۱]
snpStats[a]	مطالعات مربوط به SNP حجیم	[۴۱]

جدول ۱-۴ سایر بسته‌های استفاده شده در این کتاب. به جز موارد ذکر شده، همه بسته‌ها بر روی CRAN هستند

نام	عنوان	منبع
admixturegraph	نمودارهای ترکیبی	[۱۵۷]
CubSFS[a]	تغییر جمعیت هموار با SFS	[۲۸۸]
coalescentMCMC	تجزیه‌وتحلیل کوالسنت توسط MCMC	[۲۷۹]
fastbaps[b]	خوشه‌بندی سلسله‌مراتبی بیزی	[۱]
flashpcaR[c]	PCA با ماتریس‌های بسیار بزرگ	[۱۰۰]
Geneland[d]	ساختار جمعیت از داده‌های چند منبعی	[۸۲]
haplo.stats	تفسیر فراوانی هاپلوتایپ	[۲۰۱]
ips	رابط‌های نرم افزار فیلوژنتیک	[۳۰۱]
LEA[e]	مطالعات منظره و ارتباط اکولوژیکی	[۲۹۶]
jackalope	شبیه ساز داده‌های ژنومیک و HTS	[۱۷۴]
mmod	معیارهای مدرن تمایز جمعیت	[۲۴۷]
OutFLANK[f]	تشخیص داده‌های پرت در اسکن‌های انتخابی	[۱۵۴]
pcadapt	اسکن انتخاب با PCA	[۲۶۷]
phangorn	تخمین فیلوژنی	[۱۳۸]
phylodyn[g]	ابزارهای آماری فیلودینامیک	[]
poolSeq[h]	فراوانی‌های آللی سری زمانی	[۹۴]
poppr	ژنتیک جمعیت موجودات کلونال	[۲۷۸]
psmcr[i]	ادغام جفتی زنجیره مارکوف	[۱۶۶]
readxl	خواندن و نوشتن فایل‌های اکسل	[۱۱۲]
rehh	تست‌های مبتنی بر هموزیگوسیتی هاپلوتیپ	[۲۵۶]
rhierbaps	تحلیل بیزی ساختار جمعیت	[۵۰]
Rsamtools[e]	رابط با برنامه‌های samtools	[۳۰]
Rsubread[e]	هم‌ترازی توالی	[۱۴۸]
sangerseq[R]	بزارهایی برای داده‌های توالی‌یابی شده Sanger در R	
scrm	شبیه‌سازی ادغام زنجیرع مارکوفی	
STITCH[j]	انتساب با داده‌های HTS با پوشش کم	
tess۳r[k]	ساختار فضایی جمعیت	
vcfR	تجزیه‌وتحلیل فایل‌های VCF	

[a]https://github.com/blwaltoft/CubSFS
[b]https://github.com/gtonkinhill/fastbaps
[c]https://github.com/gabraham/flashpca/tree/master/flashpcaR
[d]https://i-pri.org/special/Biostatistics/Software/Geneland/distrib/
[e]On BioConductor
[f]https://github.com/whitlock/OutFLANK
[g]https://github.com/mdkarcher/phylodyn
[h]https://github.com/ThomasTaus/poolSeq
[i]https://github.com/emmanuelparadis/psmcr
[j]https://github.com/rwdavies/STITCH
[k]https://github.com/bcm-uga/TESS۳_encho_sen"



شکل ۱-۹ (الف) انتخاب داروینی: فراوانی افراد با فنوتیپ سودمند (رنگ خاکستری تیره) در جمعیت افزایش می‌یابد. (ب) قضیه اساسی فیشر در مورد انتخاب طبیعی: میانگین تناسب در جمعیت با نرخی متناسب با واریانس آن افزایش می‌یابد. (ج) نمونه‌ای از جهش‌های مترادف و غیرمترادف که سه ژنوتیپ مختلف را در یک ژن کدکننده پروتئین و فنوتیپ‌های مربوطه (اسیدآمینه در توالی پروتئین) نشان می‌دهد. جهش C↔ T در موقعیت اول مترادف است و تاثیری بر فنوتیپ ندارد. (د) برخی جهش‌های سودمند ممکن است دلیل بر انتخاب بر روی جایگاه‌ها باشد.

```
$ ls # this is a system command
```

نماد سریع سیستم به سیستم‌عامل (OS) یا حتی برنامه مورداستفاده برای تعامل با سیستم بستگی دارد. معمولاً C:/ در ویندوز است. اگر یک خروجی یا نتیجه خیلی طولانی باشد، با چهار نقطه کوتاه می‌شود.

```
> x <- rnorm(١e٦)
> x
   [١] ٠,٨١٧٩٩٧٧٢٧ -١,٠٠٣١٥٥٢٧٧ ١,٦٥٢٤٥٣٥٧١ ٢,٠٨٨٢٨٨٤٧٥
   [٥] ٠,٩٢٢٣٧٦٠٣٦ ٠,٩٤٦٧٤٨٥٨٠ -١,٠٢٨٩٩٦٢٨١ ١,٠٣١٢٢٩٦٥٦
....
```

نام فایل‌ها در یک نقل قول چاپ می‌شوند (به‌عنوان‌مثال، 'datafile.txt') و محتوای فایل با فونت monospace در یک قاب نشان داده می‌شود:

۱-٤ دانش لازم و سایر مطالب

این کتاب با استفاده از مفاهیم ژنتیک جمعیت، ژنومیک، بیوانفورماتیک، و آمار، کاربرد چند رشته‌ای دارد. در این فصل مطالب مقدماتی اولیه برای دو رشته اول ارائه شد. مقدمه گسترده‌تر را می‌توان در ویرایش اخیر اصول ژنتیک جمعیت توسط هارتل و کلارک [۱۰۸]

یافت. فرض بر این است که خواننده دانش پایه در مورد مفاهیم آماری زیر دارد: (کو) واریانس، احتمال، ، استنتاج بیزی و معیارهای اطلاعات. برخی از دانش بسیار ابتدایی در مورد محاسبات نیز فرض شده است (بایت، حافظه فعال، دیسک سخت) و همچنین برخی از مفاهیم اولیه در مورد حساب دیفرانسیل و انتگرال (حاصل ماتریس، مشتقات، انتگرال)، و در مورد تکنیک‌های آزمایشگاهی زیست‌شناسی مولکولی (به‌ویژه PCR و حرکت روی ژل). در نهایت، من مدل‌های آماری فریدمن [۸۱] را به‌عنوان یکی از بهترین مقدمه‌ها برای ایده‌ها و مفاهیم استنتاج آماری توصیه می‌کنم.