

پاسخنامه تمرین دوم

پرسش ۱:

این شباهتها باعث می‌شود که هنگام خواندن توالی‌ها، نرم‌افزارهای تحلیل قادر نباشند به‌طور دقیق مشخص کنند که هر قطعه خوانده‌شده به کدام بخش ژنوم تعلق دارد. به‌ویژه در تکنیک‌هایی که طول خوانش کوتاه دارند، این مسئله بیشتر دیده می‌شود. برای رفع این چالش، یکی از راه‌ها استفاده از تکنولوژی‌هایی با طول خوانش بلندتر مانند PacBio و Oxford Nanopore است که اطلاعات بیشتری از توالی‌ها ارائه می‌دهند. علاوه بر این، استفاده از الگوریتم‌های پیشرفته بیوانفورماتیکی که توانایی مدل‌سازی دقیق‌تری دارند، می‌تواند به شناسایی بهتر این بخش‌ها کمک کند. ترکیب داده‌های حاصل از تکنولوژی‌های مختلف نیز به ایجاد تصویری جامع‌تر از ژنوم و کاهش خطاهای ناشی از تکرارها منجر می‌شود.

در این روش، دو انتهای یک قطعه DNA که طول آن از قبل مشخص شده، خوانده می‌شود. این اطلاعات باعث می‌شود که نه تنها دو انتهای توالی به‌خوبی مشخص شوند، بلکه فاصله بین آن‌ها نیز داده‌ای ارزشمند برای تحلیل دقیق‌تر ژنوم باشد. در بخش‌های تکراری ژنوم، چون این نواحی بسیار شبیه به هم هستند، ابزارهای تحلیلی معمولاً نمی‌توانند تشخیص دهند که یک قطعه خوانده‌شده متعلق به کدام قسمت ژنوم است. این وضعیت که به آن "چند مسیر بودن" یا ambiguity گفته می‌شود، می‌تواند باعث شود که نتیجه نهایی نادرست باشد. اما paired-end sequencing با ارائه اطلاعات از هر دو انتهای قطعه و طول دقیق بین آن‌ها، این ابهام را کاهش می‌دهد. حتی اگر توالی‌ای تکراری باشد، این روش کمک می‌کند که مسیر صحیح تشخیص داده شود. paired-end sequencing می‌تواند تفاوت‌های کوچک یا تغییرات ظریفی که در بخش‌های تکراری وجود دارد را به‌خوبی شناسایی کند. چون اطلاعات بیشتری در اختیار نرم‌افزارهای تحلیلی قرار می‌دهد، دقت کلی فرآیند افزایش می‌یابد.



پرسش ۵: ساخت درخت‌های فیلوژنتیک

امروزه برای ساخت درخت‌های فیلوژنتیک از الگوریتم‌هایی مانند حداکثر درست‌نمایی (Maximum Likelihood)، استنتاج بیزی (Bayesian Inference) و روش اتصال همسایگی (Neighbor Joining) استفاده می‌شود و به‌طور کلی این الگوریتم‌ها امروزه خیلی پر استفاده هستند.

1. نحوه کار الگوریتم‌ها:

- **حداکثر درست‌نمایی:** در این روش احتمال مشاهده داده‌ها را تحت درخت‌های مختلف ارزیابی می‌کنیم و درختی را انتخاب می‌کنیم که بیشترین درست‌نمایی را دارد.
- **استنتاج بیزی:** بر اساس احتمالات پیشین و به‌روزرسانی آن‌ها با داده‌ها، توزیع پسین برای درخت‌ها ایجاد می‌شود (فرق اصلی آن در نظر گرفتن یک احتمال پیشین برای داده‌ها است).

2. مقایسه با اتصال همسایگی:

- **اتصال همسایگی** روش سریعی است که بر اساس فاصله‌های بین توالی‌ها درخت را می‌سازد و نسبت به داده‌های بزرگ مناسب‌تر است. اما دقت آن در مقایسه با دو روش دیگر کمتر است.
- **حداکثر درست‌نمایی و بیزی** پیچیده‌تر هستند و زمان بیشتری می‌برند، اما نتایج دقیق‌تری ارائه می‌دهند، به‌ویژه زمانی که داده‌ها دارای جهش‌های پیچیده هستند.

3. برتری‌ها:

- **بیزی:** امکان استفاده از اطلاعات پیشین و ارائه توزیع احتمالی.
- **حداکثر درست‌نمایی:** مناسب برای داده‌های پیچیده.
- **اتصال همسایگی:** سریع و ساده برای داده‌های بزرگ.

پرسش ۶: محاسبات برای الگوریتم UPGMA

برای اجرای UPGMA، ماتریس فاصله داده شده را بررسی کرده و مراحل زیر را دنبال می‌کنیم:

1. نزدیک‌ترین جفت (کمترین مقدار در ماتریس فاصله) را پیدا کرده و آن‌ها را به صورت یک خوشه ادغام کنید.
2. ماتریس فاصله جدید را با محاسبه فاصله میانگین بین خوشه جدید و سایر نقاط به‌روزرسانی کنید.
3. مراحل را تکرار کنید تا همه نقاط به یک خوشه تبدیل شوند.

به‌طور کلی ماتریس‌های هر مرحله به‌صورت زیر هستند:

• مرحله اول D و E

	A	B	C	DE	F	G
A	0					
B	5	0				
C	9	10	0			
DE	8.5	8.5	7	0		
F	7	4	10	7.5	0	
G	12	11	6	4	9	0

• مرحله دوم DE و G:

	A	BF	C	DE	G
A	0				
BF	6	0			
C	9	10	0		
DE	8.5	8	7	0	
G	12	10	6	4	0

• مرحله سوم DE و G:

	A	BF	C	DEG
A	0			
BF	6	0		
C	9	10	0	
DEG	10.25	9	6.5	0

• مرحله چهارم DEG و C:

	ABF	C	DEG
ABF	0		
C	9.5	0	
DEG	9.625	6.5	0

• مرحله پنجم A و BF:

	ABF	CDEG
ABF	0	9.5625
CDEG	9.5625	0