www.wsws.org
ලෝක සමාජවාදී වෙබ් අඩවිය
හතර වන ජාත්‍යන්තරයේ ජාත්‍යන්තර කමිටුවේ ප්‍රකාශනය

සංශ්ලේෂී (කෘතිම) ඩීඑන්ඒ අනුවක් තුල දත්ත තැන්පත් කිරීම

Data storage in synthetic DNA

බ්‍රයන් ඩයන් හා ෂේන් ෆෙරාටු විසිනි
2013 ජුලි 25

නේචර් සඟරාවේ ප්‍රසිද්ධ කර ඇති නික් ගෝල්ඩ්මන් සහ පිරිස විසින් කරන ලද නව පරීක්ෂනයකදී, ජාන විද්‍යාව හා පරිගනක විද්‍යාව සංයුක්ත කර වර්ධනය කරන ලද නව තාක්ෂනයක් මගින් ඕනෑම වර්ගයක දත්ත ඩයොක්සිරයිබෝ නියුක්ලික් ඇසිඩ් (ඩීඑන්ඒ) අනුවක් තුල විශ්වසනීය ලෙස තැන්පත් කිරීමට හැකිවේ. එමගින් මීට ඉහතදී ප්‍රායෝගික නොවූ හා පිරිවැය සමග සසඳන විට කාර්යක්ෂම නොවුනු, අවුරුදු දශක ගනනක්, ශතක ගනනක් නොව වසර දහස් ගනනක් පුරාවට දත්ත අධිලේඛනය කිරීමට අවස්ථාව ලැබෙයි. කිලෝබයිට් 739 ක දත්ත ඩීඑන්ඒ අනුවක් තුල මුල් ගොනුව 100% ක නිරවද්‍යතාවයකින් සමුද්ධරනය කිරීමට හැකිවනසේ තැන්පත් කිරීමට හැකිවේ.

මෙම සංකල්පය සකස් කිරීම සඳහා ගෝල්ඩ්මන්ගේ කන්ඩායම, ශේක්ස්පියර්ගේ පේලි 14 කින් සමන්විත සොනට්ස් කවි 154 (පාඨ ආකෘතියක් (Test format) සහ වෙනත් පාඨ ගොනුවක්, (Test file), මාටින් ලූතර් කිංගේ 1963 "මට සිහිනයක් තිබේ" යන කථාවේ ශ්‍රව්‍ය උපුටා ගැනීමක් (MP3, එම්.පී.3 ආකෘතික) සහ වර්න ඡායාරූපයක් (JPEG, ජේ.පී.ඊ.ජී ආකෘතික) ඩීඑන්ඒ අනුවක් තුල ගබඩා කලේය.

ඩීඑන්ඒ අනුවක් තුල සංඛ්‍යාංක (Digital) තොරතුරු සහිත පලමු බිඩ්ස් (තොරතුරු රැස් කිරීමේ මූලික ඒකකය) ප්‍රමානය රැස් කරන ලද්දේ 1988 දීය. එහෙත් එය ප්‍රායෝගික නොවූ, වියදම් අධික, අමාරු කටයුත්තක් විය. එයට අමතරව 2012 තරම් මෑතදී දත්ත කියවීම හා ඇතුල් කිරීම සඳහා වන වියදම හා කාලය ප්‍රායෝගික වූයේ ශතක කාල පරිමානයකට අධිලේඛනය කිරීමටය. මෑතදී අත්කරගත් වර්ධනයන් සමග එය අවුරුදු 50ක කාල පරිමානයකට කලහැකි බව අවබෝධ විය. එමෙන්ම අධිලේඛනය දශකයක් තුල කෙටි කාල පරිමානයකට පහත හෙලීම මගින් වියදමට සාපේක්ෂව සඵලතාවය ඉහල නැංවින.

පරිගනක සංඛ්‍යාංක දත්ත අධිලේඛනය, පරිගනක ක්ෂේත්‍රයේ එක් ප්‍රධාන ප්‍රශ්නයක්වේ. වර්තමාන ලෝකයේ එම දත්ත අවුරුදු 2කට වරක් දෙගුනවේ. 2020 වනවිට ලෝකයේ පරිගනක දත්ත ප්‍රමානය ගිගාබයිට් මිලියන 40 දක්වා ඉහල යන බවට ඇස්තමේන්තු කර ඇත. එහෙත් මේ සමගම, දෙනලද අවකාශයක රැස් කල හැකි දත්ත ප්‍රමානය ඉහලගොස් ඇත.

2007 වනවිට වර්ග අඟලක ක්ෂේත්‍රඵලයක් තුල රැස් කල හැකි පරිගනක දත්ත ප්‍රමානය ට්‍රිලියන 1 මට්ටම පසුකර තිබේ. ඒ අනුව දත්ත තැන්පත් කිරීම ප්‍රශ්නයක් වුවත් එය කලමනාකරනය කල හැකි තත්වයට පත්වේ.

වඩාත් දුෂ්කර ගැටළුව වනාහි, නිත්‍ය අවශ්‍යතාවයක් ලෙස පරිගනක පද්ධති ස්ථායී ලෙස නඩත්තු කර පවත්වාගෙන යාමයි. පරිගනක සංඛ්‍යාංක තොරතුරු හායනයට ලක්වීම ස්වභාවික ඵලයක් ලෙස (එනම් දෘඪතැටි වල ක්‍රියාකාරීත්වය අවසන්වීම) මෙන්ම සංඛ්‍යාංක මධ්‍යයේ නිත්‍ය වර්ධනය හා විකාශනය හේතුවෙන් යන දෙයාකාරයෙන් සිදුවේ. උදාහරනයක් ලෙස මෘදුතැටි වල (Floppy disks) වැදගත් දත්ත ගබඩාකර තිබුනත්, නූතන පරිගනක මගින් විශේෂයෙන්ම ලැප්ටොප් පරිගනක මගින් එම දත්ත නැවත ලබාගත නොහැකිය. එවිට එම දත්ත ආචය කිරීම නිෂ්ප්‍රයෝජන කටයුත්තක් වේ. මීට අමතරව දත්ත එක් උපකරනයකින් වෙනත් උපකරනයකට සමර්පනය (මාරුකිරීම) කිරීමේදී දත්ත යම් ප්‍රමානයකට හානි සිදුවේ. එම ප්‍රමානය අවුරුදු කිහිපයක් සමග සලකන විට ගනන් ගතයුතු ප්‍රමානයක් නොවුනත්, දශක ගනනක් හෝ ශතක ගනනක කාලයක් සැලකිරීමේදී සැලකිය යුතු ප්‍රමානයක් වේ.

ප්‍රධාන ගැටළුව වන සංඛ්‍යාංක පද්ධති පවත්වාගෙන යාම මෙන්ම විශාල දත්ත ප්‍රමානයක් ආචය කිරීමේදී ඩීඑන්ඒ යොදාගැනීම මගින් විසඳාගැනීමට මගක් සැලසී තිබේ. ජීවියකුගේ ජානමය දත්ත (ජිනෝමස්) එම ජීවියා විනාශ වී අවුරුදු දස දහස් ගනනක් හෝ ලක්ෂ ගනනකට පසුව පවා විශ්ලේෂනය කර තහවුරු කරනු ලැබ තිබේ. ප්‍රමිතිගත ඩීඑන්ඒ අනු ග්‍රෑම් 1ක් තුල ගිගාබයිට් බිලියන 455 ක දත්ත එහි න්‍යායිකව ඉහලම ධාරිතාවය තුල ආකේත (Encode) කලහැකිය. ජීව විද්‍යාවේදී ඉටුකල වැදගත් කාර්යය භාරය මෙන්ම ඩීඑන්ඒ අනුව තුල ඇති රසායනික සංයුතිය පරිගනක විද්‍යාවේදී නිරන්තරයෙන් (පිට පිට) දත්ත ඇතුලත් කිරීම හා එය මුල් තත්වයෙන් ලබාගැනීම සඳහා මහෝපකාරීවේ. ඩීඑන්ඒ අනුවක් තුල තොරතුරු ආචය කරනු ලබන්නේ නියුක්ලියෝටයිඩ මතට බිට්ස් ආකේත කිරීමෙනි. නියුක්ලියෝටයිඩ වනාහි ඩීඑන්ඒ සහ තවත් ඒ සමාන අනු විශේෂයක් වන ආර්.එන්.ඒ

(රයිබෝනියුක්ලික් ඇසිඩ්) ද නිර්මිත වී ඇති විශාල අනු ඒකකයි. එමෙන්ම ඇඩිනින්, ගුවනින්, ක්‍රිටොසීන්, තයිමින් (තයිමින් ආර්.එන්.ඒ තුලදී යුරසිල් මගින් විස්ථාපනයවේ.) යන එකිනෙකට වෙනස් රසායනික ආධාරක අනු හතරක් සහිත ඩීඑන්ඒ අනුව, පිරිමිඩිස් නමින් හැඳින්වෙන තනි චක්‍රයක් ලෙසත්, පියුරිස් නමින් හැඳින්වෙන යුග්ම චක්‍රයක් ලෙසත්, ව්‍යුහගතව විභේදනයව පවතී. එසේම, ද්විත්ව සර්පිලාකාර ඩීඑන්ඒ අනුවක් තත්වාකාරයෙන් තැනීම සඳහා නියුක්ලියෝටයිඩ්ස්, ඇලනීන් තයිමින් සමගත්, ගුවනින් ක්‍රිස්ටොසින් සමගත්, එකට බන්ධනගතව ඇත.

ඇලනීන් තයිමින් සමග ද ගුවනීන් ක්‍රිටොසීන් සමග ද බන්ධන ගතවෙමින් ඒවා සීනි හා පොස්පේට් අනු සමග සම්බන්ධ වෙමින් ද්විත්ව සර්පිලාකාර ඩීඑන්ඒ අනුව නිර්මානය වී ඇත.

ස්වභාවිකව පැවැතීමේදී ඩීඑන්ඒ සෛලයක අනන්‍යතාව ආකේතනය කරන අතර, එක් පරම්පරාවකින් ඊළඟ පරම්පරාවයට සෛලය පිලිබඳ විස්තර ලබාදේ. තමන්ව ජනනය වීම සහ සෛල තුල වෙනත් ප්‍රෝටීන් නිපදවීම, සෛලයක නෛසර්ගික ලක්ෂනවල ඇතිවන සැලකිය යුතු වෙනස්වීම සඳහා ප්‍රමානවත් ඉතා කුඩා වෙනස්වීමක් හෝ විකෘතිභාවයක් මෙන්ම ජීවීන් විශේෂයක එමගින් ඇතිවන මහා පරිමානයේ වෙනස් වීම්ද ඩීඑන්ඒ විසින් පාලනය කරනු ලබයි. ඩීඑන්ඒ අනුවකින් තොරතුරු ලබාගැනීමේදී එය තුල පිටපත් කිරීමේ ක්‍රියාවලියක් සිදුවිය යුතුවේ. ඒ සඳහා සර්පිලාකාර ඩීඑන්ඒ අනු, ආර්.එන්.ඒ බහු අවයවය විසින් විභේදනය කරනු ලැබේ. එහිදී ආර්.එන්.ඒ බහු අවයවය, ඩීඑන්ඒ කොටසක් සමග ආබද්ධවී අච්චුවක් සාදයි. එමගින් අනුපූරක ඩීඑන්ඒ අනුවක් නිර්මානයවේ. එයට එම් ආර්.එන්.ඒ හෙවත්, දූත ආර්.එන්.ඒ යැයි හඳුන්වනු ලැබේ. එම් ආර්.එන්.ඒ සැකසීමේ ක්‍රියාවලිය සිදුවීමෙන් අනතුරුව ඩීඑන්ඒ වෙතින් පිටපත් කරන ලද දත්ත, සෛල වල අවශ්‍යතාවය අනුව භාවිතා කිරීමට සූදානම්ව පවතී. පිටපත් කිරීමේ ක්‍රියාවලිය තුල ඇතිවන දෝෂ එනම් නියුක්ලියෝටයිඩ වැරදි තැනක ස්ථානගතවීම හෝ මැකීතිබීම වැනිදේ, ස්වයංක්‍රිය ජෛව රසායනික ක්‍රියාවලියක් මගින් අලුත්වැඩියා කරනු ලැබේ.

ගෝල්ඩ්මන් සහ පිරිස මෙයට යම් ප්‍රමානයකින් වෙනස් ක්‍රියාවලියක්, ඩීඑන්ඒ අනුවක් තුලට සංඛ්‍යාංක දත්ත ඇතුලත් කිරීම සඳහා භාවිතා කලේය. එසේ වුවත්, මූලික මූලධර්ම නොවෙනස්ව පවතී. දත්ත නියුක්ලියෝටයිඩ තුලට දත්ත ඇතුලත් කිරීම, ඒ තුල රැස් කිරීම මෙන්ම අවශ්‍ය විටක ලබාගැනීම ද සිදු කරනු ලබයි. මෑතදී ඇතිවී තිබෙන වර්ධනයන්ගේ වැදගත්කම නම් ඕනෑම වර්ගයක දත්ත තැම්පත් කිරීමට ඇති හැකියාවයි. ඊට විපරීතව එය පෙරදි විශේෂ භාවිතාවන්ට පමනක් සීමා වී පැවතියේය.

ඉතා මූලික මට්ටමේදී සෑම ගොනුවක්ම පරිඝනකයකට ලබා දෙන්නේ, 0 සහ 1 යන සංඛ්‍යා භාවිතා කරමින්, ද්වීමය (දෙකේ පාදයේ) ගනිතයෙනි. නමුත් මෙම නව සොයා ගැනීම් මගින් පරිඝනක වල අභ්‍යන්තර ක්‍රියාකාරිත්වය වඩා ඉහල මට්ටමකට එසවෙන අතර දත්ත රැස් කිරීමට හැකි ආකාර පිලිබඳ දැඩි සීමා ඉවත් වනු ඇත. පරිඝනකයකට ද්වීමය (දෙකේ පාදයේ) ගනිතයෙන් කාවද්දන ලද දත්ත 0,1 සහ 2 සහිත තුනේ පාදයට පරිවර්තනය, ඉංග්‍රිසිය ස්පාඤ්ඤ භාෂාවට පරිවර්තනය කරන්නාක් මෙන් කරනු ලබනු ඇත. එය ඊට වඩා වෙනස් විය නොහැකිය. එවිට ආකේත කරන දත්ත, ජානමය ද්‍රව්‍ය වන ඇඩිනීන්, ගුවනීන් සහ ක්‍රිස්ටොසින්, තයමින් (A"G"C"T) මගින් සටහන් කරනු ලැබේ. නිදර්ශනක් ලෙස ද්විත්ව පාදයෙන් නිරූපිත "T" අක්ෂරය ඩීඑන්ඒ තුල දක්වන්නේ "TAGAT" ලෙසය. ඒ අනුව වචන වල අකුරක් ඩීඑන්ඒ අකුරු පහකින් සමන්විත වේ.

ශේක්ස්පියර්ගේ කෙටි කවි 18 තුල ආරම්භයේ යෙදෙන "thou" වචනය "TAGATGTGTACAGACTACGC" වශයෙන් ඩීඑන්ඒ අකුරු පහක් මගින් දැක්වේ. එම්.පී.3, පීඩීඑෆ් හෝ වෙනත් ගොනු විශේෂයන්හී දත්ත මෙම ක්‍රියා සන්තතිය මගින් නිරූපනය කල හැක. එකවිට සෑදෙන අනුපිටපතක් ලබා ගැනීමට හැකිවීම, මෙම නවීනතම තාක්ෂනයෙන් ලැබෙන තවත් එක් වාසියකි. එසේම, එක් එක් වචනයක් සඳහා අසමාන ඩීඑන්ඒ ඒකක වෙනස් ආකාර හතරකින් සැදිය හැකි නිසා, දත්ත හායනයට ලක් වීම අත්‍යන්තයෙන් ම විරල වේ. මෙයට පෙර ඩීඑන්ඒ අනු තුල දත්ත රැස් කිරීමට කරන ලද පරීක්ෂන ශක්‍ය නොවුයේ, සංශ්ලේෂි පිටපත් කිරීමේදී දත්ත වලට හානි සිදුවීම නිසාය. මෙම බාධකය දැනටමත් ජය ගනු ලැබ ඇත.

ජයගතයුතු අවසන් ගැටලුව වන්නේ, ඩීඑන්ඒ තුලට දත්ත කාවැද්දීම හා එයින් දත්ත නැවත ලබා ගැනීමේ වේගය වැඩි දියුනු කිරීමයි. මේ වන තෙක් දැනට භාවිතා වන ඩිජිටල් තාක්ෂනයේ වේගයට එය ලගා වී නැත. එහෙත් දත්ත තැන්පත් කිරීමට පමනක් නොව දත්ත කාවැද්දීම හා එයින් දත්ත නැවත ලබා ගැනීමට ගත වන කාලය කෙටි කිරීමට සහ ජානමය දත්තයන්හි අනුක්‍රමනය වැඩි දියුනු කිරීමට සෑම උත්සාහයක්ම දැරෙමින් තිබේ.

ඩීඑන්ඒ අනු තුල පෘථුල දත්ත ප්‍රමානයක් අධිලේඛනය කිරීමට ඇති විශාල විභවය, ස්වභාවධර්මය පිලිබඳ මිනිසාගේ අවබෝධය වර්ධනය වීමට නව අවස්ථාවක් වන්නේය. එමෙන්ම, මනුෂ්‍යවර්ගයා ස්වභාවික ලෝකය පිලිබඳ සොයා ගැනිම් ලේඛනගත කිරීම අරම්භ කිරීමෙන් පසු පලමු වතාවට, මනුෂ්‍යවර්ගයාගේ වර්තමාන සංස්කෘතිය අනාගත පරම්පරාවන්ට තේරුම්ගත හැකි ලෙස තහවුරු කිරීම සඳහා දත්ත සංරක්ෂනය කිරීමේ සමර්ථකම අප අත්පත් කරගෙන ඇත.

© www.wsws.org