



[Título: RAPID: Inferencia filogenética en tiempo real y análisis de grupos de transmisión de COVID-19](#)

[Alexander N Moshiri CIC Perfil](#)

[NSF Award #: 2028040](#)

[Grabación de YouTube con diapositivas](#)

[November 2020 CIC Webinar Información](#)

[Transcript Editor: Shikhar Johri](#)

Transcripción:

अलेक्जेंडर मोशीरी:

स्लाइड 1

यह बढ़िया है। धन्यवाद। मुझे उम्मीद है कि लोग मेरी स्लाइड्स देखेंगे। हाँ, मैं नीमा मोशीरी हूँ। मैं यूसी सैन डिएगो में हूँ। मैं कंप्यूटर विज्ञान और इंजीनियरिंग विभाग में हूँ और मैं हाल के विकास के बारे में बात करने जा रहा हूँ जिसे हमने वायरल एमएसए कहा है।

स्लाइड 2

मूल रूप से एक मानक phylogenetic कार्यप्रवाह, यदि आप कभी भी वायरस दृश्यों के एक विकासवादी इतिहास देखा है, तो आप जरूरी बहुत अच्छी तरह से लाइन नहीं है कि शुरुआत में इन प्रारंभिक दृश्यों के साथ शुरू. और पहला कदम कई अनुक्रमों को संरेखित करना है ताकि वे पंक्तिबद्ध हों। उन्हें इस तरह संरेखित करना हमें इस बारे में बातें बताता है, आप जानते हैं, अनुक्रम होमोलॉजी क्या है, अनुक्रमों के बीच संबंध क्या हैं।

इस संरेखण का उपयोग करके, एक फाइलोजेनी का अनुमान लगाया जाता है। और फिर फाइलोजेनी यह निर्धारित करने के लिए जड़ लेता है कि सामान्य पूर्वज क्या है। और फिर अन्य डाउनस्ट्रीम विश्लेषणों का एक गुच्छा है जो करने में रुचि रखते हैं, आप जानते हैं, चीजें जैसे कि प्रकोप के क्लस्टर क्या हो रहे हैं, क्या आप जानते हैं कि हमें लगता है कि किसी बीमारी से संक्रमित होने का सबसे अधिक खतरा है? हां, कई चीजें जो आप कर सकते हैं। तो सामान्य तौर पर मेरा ध्यान इन दो चरणों पर है: बहु-अनुक्रम संरेखण और फाइलोजेनेटिक अनुमान जो वास्तव में कम्प्यूटेशनल अड़चनें हैं। और इस बात में, मैं केवल इस कदम पर ध्यान केंद्रित करूंगा: कई अनुक्रमों का संरेखण।

स्लाइड 3

तो बहु-अनुक्रम संरेखण समस्या एनपी-पूर्ण है, जो, टीएलडीआर, का अर्थ है कि हमारे पास बहुपद समय समाधान नहीं है, और परिणामस्वरूप, अनुमानित समाधानों के लिए कई हेरिस्टिक्स विकसित किए गए हैं। वे काफी सटीक हैं। कुछ उपकरण जिनसे लोग परिचित हो सकते हैं, वे हैं MUSCLE, ClustalOmega, और MAFFT। यहां कुछ सामान्य उपकरण दिए गए हैं जो उन्हें लागू करते हैं। हालांकि, यहां तक कि ये हेरिस्टिक्स आम तौर पर अनुक्रमों की संख्या के संबंध में चतुर्भुज रूप से स्केल करते हैं। इसलिए SARS-CoV-2 के मामले में, हमारे पास घातीय अनुक्रमण वृद्धि हो रही है, मेरा मतलब है, यह हमारे लिए एक बड़ी बात है। हमें अधिक से अधिक स्ट्रीम डेटा मिल रहा है, लेकिन नकारात्मक पक्ष यह है कि हमें उस स्ट्रीम डेटा का विश्लेषण करना होगा और हमारे उपकरण ठीक से स्केल नहीं करते हैं। तो अभी, यह अप्रचलित है। यह एक या दो सप्ताह पहले की बात है। अब हम लगभग 200,000 अनुक्रमों पर हैं। तो उपकरण केवल वास्तविक समय विश्लेषण की अनुमति देने के लिए पैमाने पर नहीं हैं।

स्लाइड 4

तो क्या हुआ अगर, इसके बजाय, हम पहले से जानते थे कि हमारे अनुक्रम सुपर समान होने जा रहे हैं और हमारे पास पहले से ही एक उच्च-आत्मविश्वास प्रतिनिधि संदर्भ जीनोम है जिसके खिलाफ हम उनकी तुलना कर सकते हैं? तो यहाँ, मैं संदर्भ जीनोम को शीर्ष पर एक हरे रंग की रेखा के रूप में दिखा रहा हूँ और इसके नीचे उन अन्य रंग अनुक्रमों में से प्रत्येक अनुक्रम है जिसे मैं एक दूसरे के साथ संरेखित करना चाहता हूँ। उन्हें एक-दूसरे के साथ संरेखित करने के बजाय, मैं जो कर सकता हूँ वह संदर्भ जीनोम के खिलाफ पहला अनुक्रम संरेखित करता है। संदर्भ जीनोम के साथ दूसरे अनुक्रमों को संरेखित करें। संदर्भ जीनोम के साथ स्वतंत्र रूप से अनुक्रम में से प्रत्येक को संरेखित करना जारी रखें। और फिर एंकर के रूप में संदर्भ जीनोम पदों का उपयोग करके, मैं व्यक्तिगत संरेखण को एक बहु-अनुक्रम रेखा में विलय कर सकता हूँ। तो मेरे एकाधिक अनुक्रम असाइनमेंट का पहला कॉलम देखें- पहले अनुक्रम में यह वह स्थिति है जो मेल खाती है, फिर यह स्थिति दूसरे और तीसरे में मेल खाती है, और फिर बस उन्हें एक साथ संक्षिप्त करें। मैं संदर्भ जीनोम की प्रत्येक स्थिति के लिए ऐसा कर सकता हूँ और अब मेरे पास एक से अधिक अनुक्रम रेखा है। और अच्छी बात यह है कि यह समानांतर महान है। प्रत्येक अनुक्रम को स्वतंत्र रूप से और एक साथ संदर्भ में संरेखित किया जा सकता है, और चतुर्भुज के बजाय अनुक्रमों की संख्या के साथ रैखिक रूप से तराजू किया जा सकता है।

स्लाइड 5

लेकिन मुझे बहुत जल्दी एक कदम पीछे ले जाना चाहिए और इस दृष्टिकोण के बारे में सोचना चाहिए। मेरा इनपुट एक संदर्भ जीनोम और अनुक्रमों का एक गुच्छा है जो उस संदर्भ जीनोम के समान हैं, और मेरा आउटपुट संदर्भ जीनोम के खिलाफ प्रत्येक अनुक्रम का संरेखण है। इसलिए यदि लोग लंबे समय तक पढ़े जाने वाले अनुक्रम से परिचित हैं, तो यह बिल्कुल वही कम्प्यूटेशनल समस्या है जो एक संदर्भ जीनोम को लंबे समय तक पढ़ता है। तो मेरा सवाल था: क्या मैं इस तरह के स्केलेबल संदर्भ-निर्देशित मल्टी-स्ट्रीम संरेखण को सक्षम करने के लिए मौजूदा अच्छी तरह से कार्यान्वित रीड मैपिंग टूल का लाभ उठा सकता हूँ?

स्लाइड 6

इसलिए मैंने वायरलएमएसए नामक एक उपकरण विकसित किया है जो संदर्भ-निर्देशित बहु-अनुक्रम संरेखण करने के लिए मौजूदा रीड मैपर्स के चारों ओर लपेटता है। मैं उनमें से कुछ को लपेटता हूँ, लेकिन एक विशिष्ट उपकरण है, मिनिमैप 2, जो मैं कर रहा हूँ उसके लिए अभी सोने का मानक है। इसलिए मैं सिर्फ उस विशिष्ट रीडिंग मैपर के साथ अपने टूल का उपयोग करने की सलाह देता हूँ, लेकिन मैं उनमें से

कुछ को यह दिखाने के लिए लपेटता हूं कि मैं इस टूल को स्वाभाविक रूप से विकसित कर सकता हूं क्योंकि रीडिंग मैपिंग टेक्नोलॉजीज भी विकसित होती है। तो मूल रूप से, आप बस एमएसए वायरल संदर्भ जीनोम और अनुक्रमों को संरेखित करने के लिए प्रदान करते हैं और फिर यह संदर्भ जीनोम को अनुक्रमित करने और किसी भी पूर्व-प्रसंस्करण को करने का ख्याल रखेगा जो इसे करने की आवश्यकता है, और यह रीड मैपर को कॉल करेगा और फिर विलय कर देगा एक बहु-अनुक्रम संरेखण में परिणाम।

स्लाइड 7

और यहां आप मौजूदा सर्वोत्तम अभ्यास उपकरणों के साथ इस दृष्टिकोण की तुलना देख सकते हैं। तो मेरा उपकरण नीचे नीली रेखा है और फिर अन्य दो लाइनें अन्य उपकरण हैं और आप देखते हैं कि सामान्य तौर पर, यह परिमाण के कई आदेश हैं जो लोग आम तौर पर अभी कर रहे हैं, और यह वास्तव में अच्छी तरह से तराजू है। और हमें जो अनुक्रम संरेखण मिलते हैं वे बहुत सटीक होते हैं।

स्लाइड 8

अंत में, यह उपकरण जिसे मैंने विकसित किया है, वायरल जीनोम के कई अनुक्रमों के तेजी से संरेखण की अनुमति देता है। यह खुला स्रोत है। आप इसे ऑनलाइन प्राप्त कर सकते हैं, और उम्मीद है कि यदि आप कोई वायरल परीक्षण करते हैं तो इसका उपयोग करने पर विचार करें।

स्लाइड 9

तो कुछ आभार: हेंग ली ने मिनिमैप 2 विकसित किया, जो गति का सार है और इस काम को एनएसएफ और Google द्वारा समर्थित किया गया था। और हाँ, मैं बात के लिए या बाद में सत्र में प्रश्न छोड़ दूंगा।