The COL5A1 Gene, Ultra-Marathon Running Performance, and Range of Motion

James C. Brown, Caron-Jayne Miller, Michael Posthumus, Martin P. Schwellnus, and Malcolm Collins

Purpose: Endurance running performance is a multifactorial phenotype that is strongly associated with running economy. Sit and reach range of motion (SR ROM) is negatively associated with running economy, suggesting that reduced SR ROM is advantageous for endurance running performance. The COL5A1 gene has been associated with both endurance running performance and SR ROM in separate cohorts. The aim of this study was to investigate whether COL5A1 is associated with ultra-marathon running performance and whether this relationship could be partly explained by prerace SR ROM. Methods: Seventy-two runners (52 male, 20 female) were recruited from the 56 km Two Oceans ultra-marathon and were assessed for prerace SR ROM. The cohort was genotyped for the COL5A1 BstUI restriction fragment length polymorphism, and race times were collected after the event. Results: Participants with a TT genotype (341 ± 41 min, N = 21) completed the 56 km Two Oceans ultra-marathon significantly (P = 0.014) faster than participants with TC and CC genotypes (365 ± 39 min, N = 50). The COL5A1 genotype and age accounted for 19% of performance variance. When the cohort was divided into performance and flexibility quadrants, the T allele was significantly (P = 0.044) over-represented within the fast and inflexible quadrant. Conclusion: The COL5A1 genotype was found to be significantly associated with performance in a 56 km ultra-endurance run. This study confirms previous findings and it furthers our understanding of the relationships among ROM, COL5A1, and endurance running performance. We continue to speculate that the COL5A1 gene alters muscle–tendon stiffness.

Keywords: Type V collagen, endurance, athletic performance, range of motion, running economy, genetic association
Endurance running performance is a multifactorial phenotype.\textsuperscript{1} Physiological factors that are commonly associated with endurance running ability include, but are not limited to, muscle capillary density, maximal heart rate, anthropometry, substrate utilization, and aerobic enzyme activity.\textsuperscript{1,2} Certain genetic\textsuperscript{3} and flexibility profiles\textsuperscript{4,5} have also been shown to be associated with endurance running performance.

Range of motion (ROM) is associated with running economy,\textsuperscript{6–8} a measurable marker of ultra-endurance performance.\textsuperscript{2} In elite athletes, the sit-and-reach range of motion (SR ROM) test was negatively associated with running economy.\textsuperscript{7} Similar associations have also been shown for subelite athletes. These data\textsuperscript{6–8} suggest that a reduced range of motion, as measured by the SR ROM test and other tests, is advantageous for endurance running performance.

We recently reported that the \textit{COL5A1} gene was associated with SR ROM.\textsuperscript{9,10} The \textit{COL5A1} gene encodes for the α1 chain of type V collagen, a minor fibrillar collagen that plays an essential role in fibril assembly and lateral fibril growth within connective tissues (including tendon).\textsuperscript{11} The CC genotype of the \textit{COL5A1 BstUI} restriction fragment length polymorphism (RFLP) was significantly associated with increased SR ROM.\textsuperscript{10} In contrast, Posthumus et al\textsuperscript{12} reported that the TT genotype of the \textit{COL5A1 BstUI} RFLP was associated with improved endurance running ability in the 42.2 km run of the 226 km South African Ironman triathlon. This study\textsuperscript{12} proposed that the association of the \textit{COL5A1 BstUI} RFLP and endurance running performance is mediated by changes in musculotendinous stiffness. However, a limitation of the study was that neither musculotendinous stiffness nor joint ROM were measured.

Owing to the previous association of the \textit{COL5A1 BstUI} RFLP TT genotype and improved running performance during the 226 km South African Ironman triathlon, the primary aim of this study was to investigate whether the TT genotype of the \textit{COL5A1 BstUI} RFLP was also associated with improved endurance running performance in the 56 km Two Oceans ultra-marathon.

The secondary aim of this study was to investigate whether the relationship between the \textit{COL5A1} gene and endurance running performance could be explained by prerace SR ROM. The \textit{COL5A1} gene has been previously shown to be associated with both endurance running ability and SR ROM measurements. Based on these previous findings, we hypothesize that there will be a significant over-representation of the TT genotype among those athletes with both a decreased SR ROM (inflexible) and a faster ultra-marathon finishing time.

**Methods**

**Participants**

Seventy-two Caucasian runners (52 male and 20 female) were recruited from the 2009 56 km Two Oceans ultra-marathon running race held during April in Cape Town, South Africa. The runners were of well-trained club level, with 71 of the 72 recruited athletes finishing the race. The participants of this study represent a subcohort of a larger study that investigated the association of the \textit{COL5A1 BstUI} RFLP with SR ROM measurements.\textsuperscript{10} The current study therefore represents all participants within the larger study\textsuperscript{10} (N = 325), which were recruited at and completed the Two Oceans ultra-marathon running race. Subjects were required
to complete a written informed consent before participating in this trial. SR ROM was measured at the race registration, which occurred within the three day period before the event. Sit and reach ROM was assessed by the Canadian Trunk Forward Flexion Test, which was performed as detailed in the ACSM guidelines for exercise testing with minor modification. Training data were obtained from a detailed questionnaire completed by 38 (53%) participants. Both weight and height were only self-reported in 49 (69%) of the participants. Overall race and split times were obtained from the race website (www.twooceansmarathon.org.za) after the event. Samples were genotyped anonymously. The recruitment, genotyping, and statistical analyses of this study were performed bearing the guidelines of “replicating genotype-phenotype associations” in mind. Deviations from these guidelines were noted in the manuscript. This study was approved by the Human Research Ethics Committee of the Faculty of Health Sciences within the University of Cape Town, South Africa, as well as the race organizers of the Two Oceans event.

**DNA Extraction and COL5A1 Genotyping**

Five milliliters of venous blood was obtained from each participant by venipuncture of the forearm vein and collected into an EDTA vacutainer tube. Blood samples were stored at 4°C until total DNA extraction was performed. The DNA was extracted using standard procedures as described by Lahiri and Nurnberg, and modified by Mokone et al. All participants were genotyped for the BstUI RFLP (SNP rs12722) within the 3′-untranslated region (UTR) of the COL5A1 gene (TT, TC, or CC), as previously described. Whereas a second technology was not used to verify the genotyping results, the particular RFLP had an internal digestion control that confirmed whether the enzyme had cut or not. A subset of samples (10%) were genotyped by two authors independently (JB and MP) to confirm correct genotyping of samples. Furthermore, gels were independently read and the results confirmed by two investigators.

**Statistical Analyses**

All statistical analyses were performed only on those who completed the ultramarathon. ANOVA was used to examine differences between genotype (or allele) groups and continuous data. Where appropriate, a Scheffé post hoc analysis was used to determine which of the three genotypes were significantly different from each other. Chi-squared tests were used to examine differences between genotype (or allele) groups and sex. Levene’s tests of homogeneity were performed to test for differences in homogeneity of the data. Where appropriate and hypothesis driven, the TT genotype group was compared with the combined TC and CC genotype groups. For genotype effects on running performance, both unadjusted \( P \) values and \( P \) values adjusted for weight were calculated.

In addition, the magnitude of changes in performance variables was determined on a scale of effect sizes where <0.2 = trivial, 0.21–0.6 = small, 0.61–1.2 = moderate, 1.21–2.0 = large, 2.1–4.0 = very large, and >4.0 = nearly perfect. A multivariate analysis was used to determine the model that best predicted race time with factors that were significantly associated with race time (age, weight, and COL5A1 BstUI RFLP genotype were included in the model). Hardy-Weinberg equilibrium was
calculated using Genepop version 4.0.10 (http://genepop.curtin.edu.au). Genotype frequencies were compared with previously published studies and public databases for Caucasian populations and found to be in accordance with these frequencies.

**Results**

**General Characteristics**

The \textit{COL5A1} \textit{BstUI} RFLP genotype distribution (29\% TT, \(n = 21\); 46\% TC, \(n = 33\); 25\% CC, \(n = 18\)) within this study was in Hardy-Weinberg equilibrium (\(P = 0.485\)). Age (\(P = .328\)), height (\(P = .369\)), and sex (\(P = .204\)) were similar between the three genotype groups (Table 1). However, weight (\(P = .002\)) and BMI (\(P = .027\)) were significantly different between genotype groups. Similar significant differences were observed when individuals with a TT genotype were compared with individuals with a TC or CC genotype. The average weekly distance run (km/week) during the 15 wk before the event were similar among the three genotype groups. Although the SR ROM was significantly different (\(P = .034\)) between the three genotypes, there were no significant differences when Scheffé post hoc analysis was performed on these data. In addition, there were no significant differences (\(P = .749\)) when the TT genotype was compared with the combined TC and CC genotypes.

The overall level of the race competitors, along with that of the competitors who formed part of the current study, is provided with the following summary:

- 5824 of the athletes completed the 2009 ultra-marathon within the 7 h cut-off time, of which 188 (3.2\%) finished the event in under 4 h and 924 (15.9\%) completed the race in between 4 and 5 h;
- 2400 (41.2\%) and 2312 (39.7\%) of the athletes completed the race in between 5 and 6 h, and 6 and 7 h, respectively (www.twooceansmarathon.org.za);
- the majority (\(n = 38, 53\%\)) of the 71 subjects who participated in this study finished the 56 km race in between 5 and 6 h, at an average time of 5:49 min/km (± 0:29 min, range: 4:31 to 6:47 min); and
- the remainder of the subjects completed the ultra-marathon in between 6 and 7 h, at an average time of 7:02 min/km (± 19 s, range: 4:27 to 6:40 min).

**The \textit{COL5A1} Gene and Running Performance**

Time to complete the 56 km ultra-marathon had a tendency to be different (\(P = 0.053\), weight-adjusted \(P = .046\)) among the three genotype groups (Table 1). On average, individuals with a TT genotype (341 ± 41, \(n = 21\)) were significantly (\(P = 0.014\), weight-adjusted \(P = 0.013\)) faster overall than individuals with a TC or CC genotype (365 ± 39, \(n = 50\)) (Figure 1). The magnitude of the change in performance between the individuals with a TT genotype and individuals with either a TC or CC genotype was considered “moderate” (effect size = 0.61).
<table>
<thead>
<tr>
<th></th>
<th>COL5A1 BstUI RFLP Genotype Groups</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TT</td>
<td>TC</td>
<td>CC</td>
<td>P-value&lt;sup&gt;a&lt;/sup&gt;</td>
<td>P-value&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Age (y)</td>
<td>38.5 ± 9.6 (21)</td>
<td>42.5 ± 9.3 (32)</td>
<td>40.6 ± 9.2 (18)</td>
<td>0.328</td>
<td>0.181</td>
</tr>
<tr>
<td>Sex (% males)</td>
<td>76.2 (16)</td>
<td>78.1 (25)</td>
<td>55.6 (10)</td>
<td>0.204</td>
<td>0.597</td>
</tr>
<tr>
<td>Height (m)</td>
<td>1.76 ± 0.09 (16)</td>
<td>1.79 ± 0.07 (20)</td>
<td>1.75 ± 0.07 (13)</td>
<td>0.369</td>
<td>0.554</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>69.9 ± 8.1 (17)</td>
<td>79.8 ± 12.4 (21)</td>
<td>72.5 ± 8.6 (14)</td>
<td>0.002</td>
<td>0.005</td>
</tr>
<tr>
<td>BMI (kg/m&lt;sup&gt;2&lt;/sup&gt;)</td>
<td>22.5 ± 1.7 (16)</td>
<td>24.5 ± 2.5 (20)</td>
<td>23.8 ± 2.5 (13)</td>
<td>0.027</td>
<td>0.008</td>
</tr>
<tr>
<td>Training (km/week)</td>
<td>58 ± 2 (11)</td>
<td>58 ± 14 (17)</td>
<td>49 ± 23 (9)</td>
<td>0.500</td>
<td>0.657</td>
</tr>
<tr>
<td>SR ROM (mm)</td>
<td>249 ± 91 (21)</td>
<td>231 ± 95 (32)</td>
<td>304 ± 94 (18)</td>
<td>0.034&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.749</td>
</tr>
<tr>
<td>Finishing Time (min)</td>
<td>341 ± 41 (21)</td>
<td>369 ± 34 (32)</td>
<td>357 ± 47 (18)</td>
<td>0.053</td>
<td>0.014</td>
</tr>
</tbody>
</table>

*Note.* Age, height, weight, body mass index (BMI), training (during the 15 wk period before the event), and race time are expressed as an average ± standard deviation, while sex is expressed as a frequency. The number of subjects with nonmissing data (N) is in parentheses. Age, height, weight, and training data were self-reported. BMI was calculated as kilograms per meter squared. Boldface indicates a significant difference (*P*-value < 0.05).

<sup>a</sup>TT vs TC vs CC.

<sup>b</sup>TT vs TC + CC.

<sup>c</sup>No significant differences for post hoc analyses (Scheffé).
Contributors to Variance in Race Finishing Time

For the overall running performance (56 km); age, sex, weight, and COL5A1 BstUI RFLP genotype (TT vs TC + CC) were assessed for their contribution to race time variance using a casewise multiple regression model. The best model, which accounted for 35% ($P < .001$, standard error of the estimate = 32.76) of the variance included age, weight, and COL5A1 BstUI RFLP genotype. Of these, only weight ($P = 0.018$) and genotype ($P = .038$) contributed significantly to the overall race time model. When weight was excluded from the analysis, age ($P = .018$) and genotype ($P = .047$) contributed significantly to the variance in race time (Table 2). Together with sex, these three variables accounted for 19% of the variance in performance.

Relationship Between Running Performance, COL5A1 BstUI RFLP Genotype and SR ROM

There was no correlation between overall finish time and prerace SR ROM ($r = −0.104$, $n = 71$, $P = 0.390$). However, for the purposes of this study, the genotype frequency between quadrants were analyzed (Figure 2). Quadrants were defined by the median value of SR ROM and time to complete the ultra-marathon (performance). Participants who had more ROM than the median were termed flexible,
whereas those with less ROM than the median were termed \textit{inflexible}. Similarly, those faster than the median were termed \textit{fast} and those slower than the median were termed \textit{slow} (Figure 2). All participants were divided into quadrants based on these categories: inflexible-fast (group 1), inflexible-slow (group 2), flexible-fast (group 3), and flexible-slow (group 4). To reduce the limitation of the small sample size, the T and C allele frequencies were compared between the quadrants. The T allele was significantly over-represented in participants within group 1 when compared with the remaining participants (T, 69\% vs 47\%; \( P = 0.044 \)). To determine a linear trend, groups 2 and 3 were combined. There was a significant linear trend for the T allele to be over-represented within inflexible-fast individuals, whereas the C allele was over-represented in the flexible-slow individuals (group 1,

Table 2  Multivariate analysis for the 56 km Two Oceans ultra-marathon overall finishing time (\( N = 71 \))

<table>
<thead>
<tr>
<th></th>
<th>( \beta )</th>
<th>( B )</th>
<th>( P )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y)</td>
<td>0.281</td>
<td>1.23</td>
<td>0.018</td>
</tr>
<tr>
<td>Sex</td>
<td>0.149</td>
<td>13.5</td>
<td>0.199</td>
</tr>
<tr>
<td>\textit{COL5A1} Genotype (TT vs TC + CC)</td>
<td>0.226</td>
<td>20.2</td>
<td>0.047</td>
</tr>
</tbody>
</table>

\textit{Note.} \( \beta \), partial correlation coefficient, and \( B \), parameter estimate. \( R = .436, R^2 = .191, \) standard error of the estimate = 37.73, \( P < .00257 \). Boldface indicates a significant difference (\( P \)-value < 0.05).
Brown et al.

69%, vs group 2+3, 51%, vs group 4, 36%; \( P = 0.037 \). Similar trends were found when the genotypes were compared between the four groups; however, owing to the small sample size, the significance of this analysis could not be determined (data not shown).

Discussion

The main finding of this study was that the \textit{COL5A1} BstUI RFLP TT genotype was associated with improved endurance running performance in a 56 km ultra-marathon. While there was only a trend for performances difference among the three genotypes groups, the TT genotype were significantly faster when compared with the C allele (TC and CC genotype combined). The TC and CC genotypes were grouped together based on our a priori hypothesis and small sample size. This finding confirms that of a previous study, which reported an association between the \textit{COL5A1} TT genotype and faster performance during the 42.2 km run of an Ironman triathlon.\textsuperscript{12} An additional novel finding of this study was that the T allele was significantly over-represented among inflexible-fast (group 1) individuals. This indicates a potential association between endurance running performance, \textit{COL5A1}, and SR ROM.

Endurance performance is a multifactorial phenotype resulting from a poorly understood complex interaction of environmental and genetic factors.\textsuperscript{18} The majority of genes that have been associated with endurance performance to date encode proteins involved in metabolic pathways and skeletal muscle biology.\textsuperscript{3} To our knowledge, this is the first extracellular matrix protein encoding gene that has been associated with running performance. Of further interest is that the genotype association with performance was evident in the running, but not the swimming or cycling events of the Ironman triathlon in the previous study.\textsuperscript{12} The greater amount of loading placed on the body may explain why this association is seen only with endurance running performance, and not the swimming or cycling. Furthermore, the magnitude of the effect of \textit{COL5A1} on endurance running performance was calculated as being “moderate” in this study. Owing to the polygenic nature of the endurance phenotype,\textsuperscript{3} it is highly unlikely that a single gene would have any greater magnitude of effect. Based on the current study design and the limited available training data, it was not possible to determine whether this gene-performance interaction has a direct effect on running performance during the event or enables athletes to train harder in preparation for the race. Both mechanisms are plausible and further research is required to elucidate this question. As with other biological systems, connective tissue is encoded by multiple genes. It is therefore likely that other extracellular matrix encoding genes need to be considered for their contribution to the performance phenotype.

Together with the \textit{COL5A1} genotype (TT vs TC + CC), age and sex accounted for 19% of the variance in running performance for the ultra-marathon. Although the contribution of age and sex to endurance running performance was previously described, the contribution of \textit{COL5A1} genotype to performance is novel and worth exploring in future studies.

The \textit{COL5A1} genotype alone accounted for approximately 7% of the variance in performance and contributed significantly to the model (\( P = .027 \)). The fact that weight was significantly associated with the \textit{COL5A1} BstUI RFLP genotype
in this cohort cannot be explained, and it should be reiterated that not all subjects reported their height and weight. This finding may therefore be spurious and should be interpreted with caution owing to the small sample size. In addition, there is no evidence for a COL5A1 BstUI RFLP genotype-weight association from the published literature of several larger diverse cohorts.10,12,16,19,20

We recently reported10 that the CC genotype was associated with greater SR ROM in a larger cohort consisting of 325 subjects. The participants of the current study represent part of a larger study that investigated the association of the COL5A1 BstUI RFLP with SR ROM measurements. The current study included only the participants who completed the 2009 Two Oceans 56 km ultra-marathon, and therefore it was not an objective of this particular study to investigate the relationship between SR ROM and the COL5A1 gene. However, this subcohort enabled us to investigate the previously proposed relationship between running performance, SR ROM, and COL5A1 genotype.12

In the current study, we did not find a direct association between SR ROM measurements (flexibility) and time to complete the 56 km ultra-marathon race. We believe that Figure 3 explains this lack of association due to both ROM and endurance running performance being complex multifactorial phenotypes.1,21 Although there are associated factors that are common to both endurance running performance and joint ROM (eg, age, sex, and the COL5A1 BstUI genotype),9,12 there are also unique factors, which are associated with only one or the other phenotype (eg, slow-twitch muscle fiber proportion for endurance performance and limb dominance for ROM) (Figure 3).22,23 We propose that muscle-tendon stiffness—an additional intrinsic factor common to both running performance and SR ROM24,25—is the most plausible biological explanation for the findings of this study and previously published associations with the COL5A1 gene.9,10,12,20 In support of this proposed mechanism, COL5A1 haplo-insufficiency in mice (± mutants) leads to a significantly different connective tissue elasticity modulus (stiffness).26 This finding therefore supports a role of COL5A1 in modulating the biomechanical properties of muscle-tendon units. Moreover, SR ROM and running performance are both determined by, among other factors, the biomechanical properties of the musculoskeletal system. Increased muscle-tendon stiffness is associated with a reduced ROM24 and increased running performance.25 It is, therefore, not surprising that in our study, there was an over-representation of the T allele within the “inflexible-fast” athletes. Conversely, there was an over-representation of the C allele within the “flexible-slow” athletes. Because of the small sample sizes, this finding should be interpreted with caution and repeated in a larger cohort.

The main limitations of this study were the small sample sizes within genotype groups and that not all participants had height and weight measurements recorded. The fact that weight was significantly associated with genotype emphasizes the latter limitation. Nevertheless, running performance between the genotype groups remained significant after adjusting for the available weight measurements. Furthermore, the COL5A1 genotype significantly contributed to the multiple regression model irrespective of whether weight was included or excluded from the analysis. Future biomechanical studies are required to further explore the relationships detailed in this study. Direct measurements of muscle-tendon stiffness are required to gain a clearer insight into the mechanism of the association between an extracellular matrix encoding gene and endurance running performance.
In conclusion, the **COL5A1** TT genotype was associated with improved endurance running performance in a cohort of ultra-endurance athletes. The sit-and-reach ROM of this genotype did not explain this association with performance, but the T allele was significantly over-represented in the inflexible-fast athletes within this study. Larger cohort studies are required before the relationship between endurance running performance and the **COL5A1 BstUI RFLP** can be fully understood.

**Practical Applications**

The study confirms a previous novel finding\textsuperscript{12} that the BstUI RFLP TT genotype of **COL5A1** is associated with faster endurance running performance. While there was no direct relationship between prerace SR ROM and running performance, the T allele was significantly over-represented within the fast-inflexible athletes.
Weight, height, and training data were not available for all the athletes; nevertheless, age and \textit{COL5A1} genotype accounted for a substantial proportion (19%) of running performance in this cohort. As result, this study emphasizes the importance of considering genes that encode for connective tissue proteins within musculoskeletal soft tissues when examining endurance running performance. Furthermore, the over-representation of the \textit{COL5A1} T allele in the fast-inflexible athletes supports our hypothesis that there is a relationship between the \textit{COL5A1} gene, joint ROM, and endurance running performance. This relationship should be further investigated within a larger cohort, using direct measures of muscle-tendon stiffness and running economy.

\textbf{Acknowledgments}

This work was supported, in part, with funds from the South African National Research Foundation (grant number IFR2009021600020 to MC), the South African Medical Research Council and the University of Cape Town. Brown was supported by postgraduate funding from the University of Cape Town, as well as funding from the Department of Science and Technology and the National Research Foundation in South Africa. The authors would also like to thank all the staff and students within the UCT/MRC Research Unit for Exercise Science and Sports Medicine who assisted with data collection during the Two Oceans ultra-marathon.

\textbf{References}