The *COL5A1* Gene: A Novel Marker of Endurance Running Performance

MICHAEL POSTHUMUS[1], MARTIN P. SCHWELLNUS[1], and MALCOLM COLLINS[1,2]

[1]*UCT/MRC Research Unit for Exercise Science and Sports Medicine of the Department of Human Biology, Faculty of Health Sciences, University of Cape Town, Cape Town, SOUTH AFRICA; and* [2]*The South African Medical Research Council, Cape Town, SOUTH AFRICA*

ABSTRACT

POSTHUMUS, M., M. P. SCHWELLNUS, and M. COLLINS. The *COL5A1* Gene: A Novel Marker of Endurance Running Performance. *Med. Sci. Sports Exerc.*, Vol. 43, No. 4, pp. 584–589, 2011. **Background**: Running economy, a key component of endurance ability, has been shown to be associated with flexibility. Increased stiffness (inflexibility) may improve running economy and therefore endurance running ability. The *COL5A1* gene, which encodes the α1-chain of type V collagen, was found to associate with measures of flexibility. Type V collagen is a quantitatively minor fibrillar collagen, which is believed to regulate fibrillogenesis within tendons and other connective tissue. **Purpose**: The aim of this study was therefore to determine whether the *COL5A1* gene is associated with endurance performance. **Methods**: Three hundred thirteen Caucasian male participants who completed either the 2006 or the 2007 226-km South African Ironman triathlon (3.8-km swim, 180-km bike, and 42.2-km run) participated in this study. All participants were genotyped for the *COL5A1 Bst*UI restriction fragment length polymorphism (RFLP). **Results**: The *COL5A1 Bst*UI RFLP was significantly associated with time to complete the running component of the triathlon. Participants with a TT genotype completed the running component of the race significantly faster than individuals with a CC genotype ($P = 0.019$; mean ± SD: TT = 294.2 ± 52.1 min, CC = 307.4 ± 48.6 min). In addition, there was a significant linear trend ($P = 0.020$) in the CC genotype distribution when the run times were divided into the fastest (13%), middle (17%), and slowest (25%) tertiles. There were no significant genotype differences for time to complete the swim, the bike, or the overall race. *COL5A1 Bst*UI RFLP, body mass index, age, and 15 wk of running training history predicted 30% of the variance in running performance. **Conclusion**: This is the first study to identify the *COL5A1 Bst*UI RFLP as a marker for endurance running performance. Further studies are required to replicate these findings. **Key Words:** TYPE V COLLAGEN, IRONMAN, TRIATHLON

BASIC SCIENCES

It is generally accepted that athletic ability is determined by both environmental and genetic factors (16). Various genetic loci and markers have been reported to be associated with physical performance or health-related phenotypes (4). More specifically, the 2006–2007 gene map reported 47 loci that associate with endurance phenotypes (4). Because multiple genetic markers have been shown to be associated with endurance performance, it remains possible that selected variants within previously uninvestigated candidate genes may further explain the genetic contribution to an optimal endurance profile.

Running economy, $\dot{V}O_{2max}$, and lactate threshold are key components that determine endurance performance (14). Measures of flexibility have been shown to be associated with running economy (9,10,13). A decreased general lower body flexibility score (increased tightness) was associated with a lower steady state $\dot{V}O_2$ during treadmill running and walking (10). Moreover, increased range of motion (ROM) measurements, such as increased ankle dorsiflexion and standing hip rotation (9), and increased sit-and-reach measurements (13), were also associated with a decrease in running economy. These data suggest that inflexibility improves running performance, possibly through enhancing the storage and return of energy and minimizing the need for muscle-stabilizing activity (9).

Like endurance performance, human ROM is influenced by several factors, which include environment (such as flexibility training and physical activity) (8) and genetic factors (1,6,12,17). However, the heritability component of ROM has been shown to be high (in the range of 64%–72%) (1,12,17). More specifically, the *COL5A1* gene *Bst*UI restriction fragment length polymorphism (RFLP) has recently been shown to be associated with passive straight leg raise and/or a sit-and-reach measurement (5,6). Mutations within the *COL5A1* gene, which encodes the α1 chain of type V collagen, causes several classic forms of the Ehlers–Danlos syndrome, for which joint hypermobility is a key clinical feature (23). Type V collagen is a quantitatively minor fibrillar collagen, which is believed to initiate fibril assembly and regulate lateral fibril growth within tendons and ligaments (3,23).

It has been previously shown that ROM measurements are associated with endurance performance (i.e., running

Address for correspondence: Malcolm Collins, Ph.D., UCT/MRC Research Unit for Exercise Science and Sports Medicine, P.O. Box 115, Newlands, 7725, South Africa; E-mail: malcolm.collins@uct.ac.za.
Submitted for publication May 2010.
Accepted for publication July 2010.

0195-9131/11/4304-0584/0
MEDICINE & SCIENCE IN SPORTS & EXERCISE®

DOI: 10.1249/MSS.0b013e3181f34f4d

economy) (9,10,13) and that ROM measurements have been associated with the *COL5A1 Bst*UI RFLP (5,6). Therefore, the objective of this study was to determine whether the *COL5A1 Bst*UI RFLP was associated with endurance performance during the 2006 and 2007 South African Ironman triathlons. More specifically, the primary aim of this study was to determine whether the *COL5A1 Bst*UI RFLP genotypes (TT, TC, or CC) are associated with the finishing time for the 3.8-km swim, the 180-km bike, the 42.2-km run as well as the overall ironman triathlons.

METHODS

Participants. Participants were recruited from the 1925 male finishers of either the 2006 (813 finishers) or the 2007 (1112 finishers) 226-km South African Ironman triathlons. Both triathlons were multiphased ultraendurance events consisting consecutively of a 3.8-km swim, a 180-km bike, and a 42.2-km run. Before the event, each competitor was sent a detailed explanation of the study and invited to participate. At race registration, those triathletes who agreed to participate in the study completed an informed written consent form as well as personal particulars and a training-related questionnaire. Three hundred thirteen consenting male triathletes (251 and 62 from the 2006 and 2007 South African Ironman triathlons, respectively) were included in this study. Data from the 2006 event were included in the analysis if the same athlete consented to participate during both the 2006 and the 2007 ironman triathlons. To avoid possible population stratification, only Caucasian men were included in this study. Ethical approval for this study was granted by the Research Ethics Committee of the Faculty of Health Sciences within the University of Cape Town, South Africa.

DNA extraction and genotyping. Approximately 4.5 mL of venous blood was obtained from each participant by venipuncture of a forearm vein and collected into an EDTA vacutainer tube. Blood samples were stored at 4°C until total DNA extraction. DNA was extracted using the procedure described by Lahiri and Nurnberger (15) and modified by Mokone et al. (18). A 667-bp fragment containing the *Bst*UI RFLP (SNP rs12722) within the 3′-UTR of the *COL5A1* gene was PCR amplified as described by Greenspan and Pasquinelli (11) and modified by Mokone et al. (18). The C and the T alleles of the *COL5A1 Bst*UI polymorphism were identified by digesting the PCR products with the restriction endonuclease, *Bst*UI, as previously described (18). The resultant fragments were separated together with a 100-bp DNA ladder of known size markers (Promega Corporation, Madison, WI) and the SYBER® Gold nucleic acid gel stain (Invitrogen Molecular Probes™, Eugene, Oregon) on 6% nondenaturing polyacrylamide gels. The gels were photographed under UV light using a Uvitec photodocumentation system (Uvitec Ltd., Cambridge, UK), and genotypes were determined on the basis of the sizes of the DNA fragments.

Statistical analysis. Data were analyzed using Statistica Version 9.0 (Statsoft, Inc., Tulsa, OK) and Graphpad InStat Version 3 (Graphpad Software, San Diego, CA) statistical programs. Any significant differences in characteristics between the three *COL5A1 Bst*UI RFLP genotype groups were tested by a one-way ANOVA. When the overall F value was significant, a Tukey's honest significance *post hoc* test was used to determine the specific differences. A chi-squared (χ^2) analysis or a Fisher's exact test was used to analyze any differences in the genotype and allele frequencies as well as other categorical data between the groups. Bivariate correlations were used to determine the relationship between split times and overall race and physiological and training parameters. Age, body mass index (BMI), training (for each specific component or overall race, as appropriate), and the *COL5A1 Bst*UI RFLP genotype data were used in a multivariate analysis by a forward stepwise regression. The four independent multivariate analyses determined which variables (model) best predicted performance during each of the components (i.e., swim time, bike time, and run time) as well as overall time for the 2006 and 2007 South African Ironman triathlons. Significance was accepted when $P < 0.05$. The Hardy–Weinberg equilibrium was established using the program Genepop Web version 3.4 (http://genepop.curtin.edu.au/).

RESULTS

Participant characteristics. The mean ± SD values of the finishing times for the swim, bike, run, and overall ironman triathlons were 90 ± 16, 401 ± 40, 294 ± 52, and

TABLE 1. The physical characteristics and performance during the ironman triathlons of all participants as well the respective *COL5A1 Bst*UI RFLP genotype groups.

		The *COL5A1 Bst*UI RFLP Genotype Groups			
	All Participants (*N* = 313)	**TT (*n* = 115)**	**TC (*n* = 141)**	**CC (*n* = 57)**	***P***
Physical characteristics					
Age (yr)	38.7 ± 8.3 (246)	38.8 ± 8.4 (95)	38.5 ± 8.8 (105)	38.8 ± 7.1 (46)	0.969
Weight (kg)	78.7 ± 9.7 (256)	79.7 ± 9.8 (101)	77.5 ± 9.7 (110)	79.4 ± 9.3 (45)	0.210
Height (cm)	180.7 ± 6.8 (243)	180.6 ± 6.4 (97)	180.0 ± 6.9 (106)	180.4 ± 7.1 (40)	0.236
BMI ($kg \cdot m^{-2}$)	24.1 ± 2.5 (242)	24.2 ± 2.5 (97)	23.9 ± 2.6 (105)	24.3 ± 1.95 (40)	0.624
Country of birth (% SA born)	55.8 (249)	52.0 (98)	55.1 (107)	65.9 (44)	0.301
Ironman performance					
3.8-km swim (min)	89.5 ± 16.2 (311)	89.8 ± 15.8 (115)	89.5 ± 17.5 (139)	89.0 ± 13.5 (57)	0.957
180-km cycle (min)	401.3 ± 40.4 (310)	398 ± 38.8 (114)	400.5 ± 42.6 (139)	410.0 ± 37.4 (57)	0.178
42.2-km run (min)	294.2 ± 52.1 (311)	284.7 ± 47.8 (114)	296 ± 55.7 (140)	307.4 ± 48.6 (57)	0.021
Overall finish time (min)	788.3 ± 95.5 (313)	774.4 ± 90.2 (115)	791.4 ± 102.3 (141)	808.8 ± 85.1 (57)	0.074

Country of birth is presented as a frequency (%); the remaining variables are expressed as a mean ± SD. The number of participants with nonmissing or valid data (*n*) is in parentheses.
BMI, body mass index.

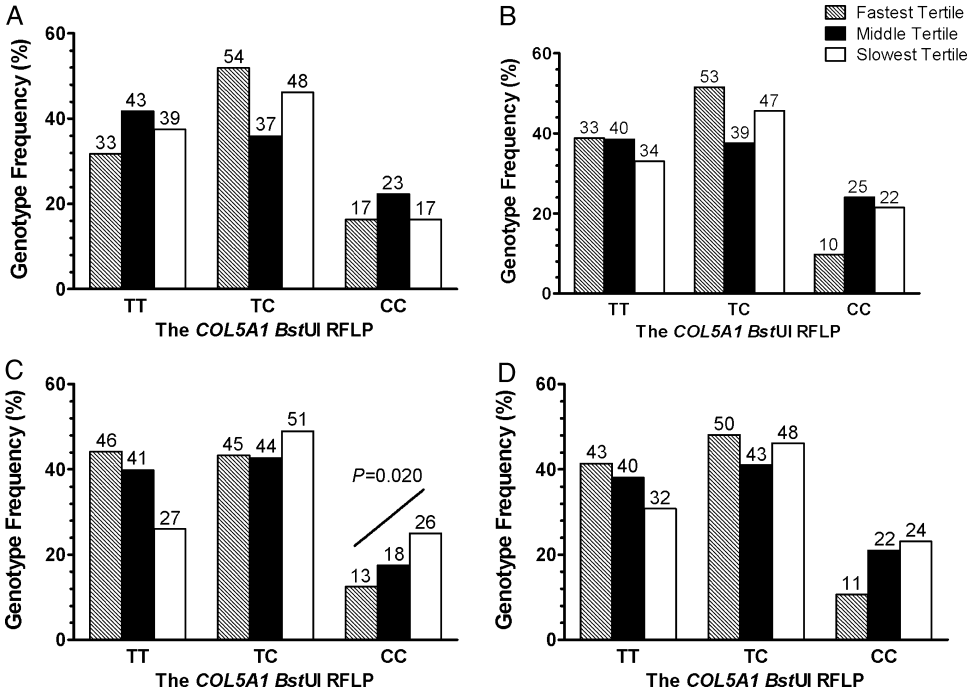

FIGURE 1—The *COL5A1 Bst*UI RFLP genotype frequencies within the tertile groups of the finishing times of the 3.8-km swim (A), 180-km bike (B), 42.2-km run (C), and the overall ironman triathlons (D). The number of participants with nonmissing or valid data corresponding to the particular genotype and tertile group is indicated above each bar.

788 ± 96 min, respectively (Table 1). These times are representative of all finishers of the 2006 and 2007 ironman triathlons (data not shown). All 313 participants were genotyped for the *COL5A1 Bst*UI RFLP genotype. One hundred and fifteen participants (37%) had a TT genotype, 141 participants (45%) had a TC genotype, and 57 (18%) had a CC genotype. The *COL5A1 Bst*UI RFLP was in the Hardy–Weinberg equilibrium (P = 0.246). As shown in Table 1, there were no significant differences in age, height, weight, BMI, or country of birth between the three genotype groups.

The *COL5A1 Bst*UI RFLP and endurance performance. There was a significant difference in the finishing time of the run when the triathletes were divided into the three *COL5A1 Bst*UI RFLP genotype groups (Table 1). The mean run times of the triathletes with a TT genotype were significantly faster than those with a CC genotype (P = 0.019). However, there were no significant differences for the finishing times of the swim or the bike as well as the overall time to complete the ironman triathlons. Similar results were obtained when only the 2006 data were analyzed separately (data not shown).

In addition, all participants were divided into tertile groups (fastest tertile, middle tertile, and slowest tertile) according to their finishing time for each discipline (swim, bike, and run) as well as the overall finishing times (Fig. 1). There was a significant increase in the CC genotype frequency within the run tertile groups (P value for linear trend = 0.020), with the CC genotype frequency of 13% (13 of 104), 17% (18 of 103), and 25% (26 of 104) in the fastest, middle, and slowest tertiles, respectively. The genotype frequency distribution among the tertile groups were not significantly different for the swim (P = 0.216), the bike (P = 0.052), and the overall ironman triathlons (P = 0.116). There were also no significant differences in the mean training volumes ($h \cdot wk^{-1}$ or $km \cdot wk^{-1}$) during the 15 wk before the events between the three genotype groups (Table 2). Furthermore, when only participants

TABLE 2. Self-reported training history (time and distance) 15 wk before the ironman triathlons within the *COL5A1 Bst*UI RFLP genotype groups.

		The *COL5A1 Bst*UI RFLP Genotype Groups			
	All Participants	TT	TC	CC	*P*
Training volume 15 wk before the ironman race					
Swimming ($h \cdot wk^{-1}$)	3.1 ± 1.6 (251)	2.9 ± 1.1 (97)	3.1 ± 2.0 (109)	3.1 ± 2.0 (45)	0.585
Bike ($h \cdot wk^{-1}$)	8.8 ± 11.6 (232)	3.2 ± 2.9 (90)	9.6 ± 17.4 (100)	8.2 ± 2.9 (42)	0.668
Running ($h \cdot wk^{-1}$)	4.8 ± 2.9 (238)	4.8 ± 1.9 (92)	4.8 ± 3.9 (101)	4.7 ± 1.7 (45)	0.386
Total ($h \cdot wk^{-1}$)	15.6 ± 4.4 (238)	15.8 ± 4.7 (87)	15.1 ± 4.1 (93)	16.1 ± 4.5 (41)	0.386
Swimming ($km \cdot wk^{-1}$)	6.3 ± 3.0 (252)	6.3 ± 2.9 (99)	6.3 ± 3.2 (108)	6.5 ± 2.7 (45)	0.943
Bike ($km \cdot wk^{-1}$)	233.1 ± 82.4 (230)	237.5 ± 91.0 (92)	215.8 ± 75.9 (97)	207.9 ± 73.6 (41)	0.083
Running ($km \cdot wk^{-1}$)	46.5 ± 18.1 (247)	47.6 ± 17.4 (99)	45.3 ± 17.7 (104)	46.7 ± 20.7 (44)	0.690
Total ($km \cdot wk^{-1}$)	236.9 ± 82.9 (211)	252.1 ± 91.2 (86)	230.0 ± 76.8 (87)	218.0 ± 72.2 (38)	0.065

The variables are expressed as a mean ± SD. The number of participants (n) with nonmissing or valid data is in parentheses.

TABLE 3. Multivariate analysis for time to complete the swim (swim time), the bike (bike time), the run (run time), and the overall ironman triathlons (overall time).

	β	B	P
Swim time (min)			
Age (yr)	0.308	0.593	<0.001
15 wk swim training (km·wk^{-1})	0.245	−1.228	<0.001
BMI (kg·m^{-2})	0.123	0.776	0.045
Bike time (min)			
15 wk cycle training (km·wk^{-1})	−0.400	−0.18	<0.001
BMI (kg·m^{-2})	0.261	3.87	<0.001
Age (yr)	0.153	0.76	0.011
Run time (min)			
BMI (kg·m^{-2})	0.496	10.42	<0.001
15 wk run training (km·wk^{-1})	−0.192	−0.57	<0.001
*COL5A1 Bst*UI RFLP genotype pair	0.122	13.11	0.018
Overall time (min)			
BMI (kg·m^{-2})	0.401	14.34	<0.001
15 wk total training (km·wk^{-1})	−0.387	−0.416	<0.001
Age (yr)	0.118	1.38	0.046

For the swim time: $R = 0.454$, adjusted $R^2 = 0.195$, SEE = 13.8, $P < 0.0001$. For the bike time: $R = 0.533$, adjusted $R^2 = 0.1273$, SEE = 32.4 $P < 0.0001$. For the run time: $R = 0.555$, adjusted $R^2 = 0.296$, SEE = 44.4, $P < 0.0001$. For the overall time: $R = 0.614$, adjusted $R^2 = 0.368$, SEE = 71.9, $P < 0.0001$.
β, partial correlation coefficient; *B*, parameter estimate.

with complete training history data were analyzed, the run time remained significantly different between the three genotype group (data not shown).

Multivariate analysis for the determination of performance. The athlete's age ($r = 0.225$, $P < 0.001$), BMI ($r = 0.462$, $P < 0.001$), and overall training load in the last 15 wk (km·wk^{-1}) ($r = 0.407$, $P < 0.001$) were all significantly correlated with overall race time. A bivariate analysis was therefore used to describe the relationship between the split times of the swim, bike, and run as well as the overall time to complete the ironman triathlons and the selected variables. The participants' *COL5A1 Bst*UI RFLP genotype pairs (TT vs TC + CC), age, BMI, and discipline-specific training load in the last 15 wk (km·wk^{-1}) were incorporated in the model to determine performance for the swim, bike, run, and overall triathlon (Table 3). The variables age, last 15 wk of swim training (km·wk^{-1}), and BMI predicted 20% of the variance ($P < 0.0001$, SEE = 13.8) in the time to complete the swim. For the bike, the variables last 15 wk of bike training (km·wk^{-1}), BMI, and age predicted 27% of the variance ($P < 0.0001$, SEE = 32.4). All four variables, including the *COL5A1 Bst*UI RFLP genotype pair (TT vs TC + CC), predicted 30% of the variance ($P < 0.0001$, SEE = 44.4) in the time to complete the run. The variables BMI, total training in last 15 wk (km·wk^{-1}), and age predicted 37% of the variance ($P < 0.0001$, SEE = 71.9) in the overall time to complete the ironman triathlons.

DISCUSSION

This is the first study to identify the *COL5A1 Bst*UI RFLP (a gene that encodes a structural protein of the extracellular matrix) as a marker for endurance running ability. The time to complete the run but not the swim, bike, or overall finishing time of the 2006 and 2007 ironman triathlons was associated with the *COL5A1 Bst*UI RFLP. The primary finding of this study was that male triathletes with a TT genotype completed the running component during the 2006 and 2007 South African Ironman triathlons significantly faster than individuals with a CC genotype. In addition, there was a significant linear trend in the CC genotype frequencies of the tertile groups for the running component. The fastest tertile of triathletes had the lowest CC genotype frequency. Moreover, *COL5A1 Bst*UI RFLP, BMI, age, and running training during the last 15 wk predicted 30% of the variance in the time to complete the run.

Although the findings of the study do not exclude the possibility that the *COL5A1 Bst*UI RFLP could be in linkage disequilibrium with a causative variant(s) within a neighboring gene, it is possible that variants within the *COL5A1* gene are directly associated with running performance. If *COL5A1* is directly involved, the exact mechanism by which the *COL5A1 Bst*UI RFLP or the causative variant within the gene affects endurance running ability remains largely unknown. However, it is possible that the *COL5A1 Bst*UI RFLP and the running performance association are mediated by changes in musculotendinous flexibility. Further studies, however, are required to test this proposed relationship.

A recent study, however, has shown that the *COL5A1 Bst*UI RFLP was associated with ROM measurements (5). In this study, consisting of 325 apparently healthy and physically active participants, those with a CC genotype were "protected" against the commonly reported age-related decline in flexibility measurements (5). Although the current study did not measure musculoskeletal flexibility, previous studies have shown that measures of flexibility are associated with running performance (9,10,13). A study by Gleim et al. (10) was first to report this relationship. They assigned their participants into three groups, namely, the "tightest," "normal," and "loosest" individuals, as scored by general body flexibility tests. The "tightest" group of individuals had a lower steady-state $\dot{V}O_2$ during walking and jogging (10). This finding has since been confirmed in male subelite marathon runners (9) and international-standard distance runners (13). Among the subelite marathon runners, dorsiflexion and standing hip rotation were significantly associated with the aerobic demand of running, indicating that stiffer athletes have a more optimal endurance profile (9). In agreement, stiffer international-level distance runners, as measured by the sit-and-reach test, were also more economical (13).

The *COL5A1* gene encodes the pro-α1 chain of type V collagen, the rate-limiting component of the of type V collagen trimer assembly (24). Heterotypic collagen I/V interactions are believed to regulate the fibril diameter and fibril number *in vitro* (23). The functional loss of one allele of *COL5A1* (haploinsufficient) accounts for 40% of all cases with *classic* Ehlers–Danlos syndrome (EDS) (21,25). EDS is a Mendelian genetic disorder characterized by joint laxity, fragile hyperextensible skin, and various other manifestations of connective tissue weakness (2). In a murine model of EDS (*col5a1* haploinsufficient mice), the fibril number and the collagen content are reduced, whereas the

fibril diameter has a broader distribution and demonstrates an asymmetric distribution with an increase in the larger diameter fibrils (24,25). Further research using molecular techniques is required to establish if the *COL5A1 Bst*UI RFLP affects the *COL5A1* expression and therefore the rate of type V collagen production. Although it is not known if the *COL5A1 Bst*UI RFLP is functional, it is tempting to speculate that this genetic variant affects collagen assembly and regulation. Because of the classical presentations of the Ehlers–Danlos syndrome, it is accepted that collagen dysregulation may affect flexibility and connective tissue stiffness (2,24,25).

Moreover, previous studies have also shown that the TT genotype of the *COL5A1 Bst*UI RFLP was significantly overrepresented among individuals with Achilles tendinopathy in two independent populations (18,22) as well as female participants with anterior cruciate ligament ruptures (19) when compared with their respective control groups. We therefore also speculate that this "increased risk" phenotype may be the result of increased musculotendinous stiffness and decreased ROM.

The current study has only investigated the effect of a single genetic variant on endurance performance. It is important to note that endurance ability is a polygenic phenotype (4,16). Various other genetic variants have been associated with endurance ability (reviewed by Bray et al. (4)) and more specifically ironman triathlon performance (7,20). In previous studies from our laboratory, the angiotensin converting enzyme (7), the nitric oxide synthase 3 (20), and the bradykinin beta 2 receptor (20) genes have been associated with performance during the 2000 and/or 2001 South African Ironman triathlon.

Because of the multifactorial nature of endurance performance, it is important to consider the possible effects of additional nongenetic factors (16). For this reason, the current study is strengthened by the fact that all three genotype groups were matched for training distance and training duration of each component of the triathlon. As previously mentioned, it is however a limitation that other factors, such as flexibility training, were not measured. Secondly, in our study, measures of flexibility were not measured. The initial design of this study was to investigate if the *COL5A1* gene was associated with endurance ability, and it was therefore beyond the scope of this field research to measure flexibility in each participant. On the basis of these findings, future research should, however, carefully plan studies so that sufficient variables are measured to enhance our understanding of the link between genetics, flexibility, athletic performance, and injury risk. Although we have speculated that the *COL5A1* gene effects running economy through mechanisms previously discussed, we cannot exclude the possibility that performance in this unique endurance event is determined by a unique set of physiological and psychological demands.

In conclusion, a variant within the *COL5A1* gene was associated with increased running ability during the ironman triathlon in this study. Further studies are required to confirm this novel finding. We speculate that the TT genotype of the *COL5A1 Bst*UI RFLP results in an increased musculotendinous stiffness and therefore greater running economy. However, this proposed link should be tested in future studies.

This study was supported in part by funds from the National Research Foundation (NRF) of South Africa (grant Nos. FA2005 021700015 and FA2007032700010), the South African Medical Research Council, and the University of Cape Town.

The authors report no conflict of interest.

The results of the present study do not constitute endorsement by the American College of Sports Medicine.

REFERENCES

1. Battie MC, Levalahti E, Videman T, Burton K, Kaprio J. Heritability of lumbar flexibility and the role of disc degeneration and body weight. *J Appl Physiol*. 2008;104(2):379–85.
2. Beighton P, De PA, Steinmann B, Tsipouras P, Wenstrup RJ. Ehlers–Danlos syndromes: revised nosology, Villefranche, 1997. Ehlers–Danlos National Foundation (USA) and Ehlers–Danlos Support Group (UK). *Am J Med Genet*. 1998;77(1):31–7.
3. Birk DE, Fitch JM, Babiarz JP, Doane KJ, Linsenmayer TF. Collagen fibrillogenesis *in vitro*: interaction of types I and V collagen regulates fibril diameter. *J Cell Sci*. 1990;95:649–57.
4. Bray MS, Hagberg JM, Perusse L, et al. The human gene map for performance and health-related fitness phenotypes: the 2006–2007 update. *Med Sci Sports Exerc*. 2009;41(1):35–73.
5. Brown JC, Miller C-J, Schwellnus MP, Collins M. Range of motion measurements diverge with increasing age for COL5A1 genotypes. *Scand J Med Sci Sports*. in press.
6. Collins M, Mokone GG, September AV, Van der ML, Schwellnus MP. The *COL5A1* genotype is associated with range of motion measurements. *Scand J Med Sci Sports*. 2009;19(6):803–10.
7. Collins M, Xenophontos SL, Cariolou MA, et al. The ACE gene and endurance performance during the South African Ironman Triathlons. *Med Sci Sports Exerc*. 2004;36(8):1314–20.
8. Corbin CB. Flexibility. *Clin Sports Med*. 1984;3(1):101–17.
9. Craib MW, Mitchell VA, Fields KB, Cooper TR, Hopewell R, Morgan DW. The association between flexibility and running economy in sub-elite male distance runners. *Med Sci Sports Exerc*. 1996;28(6):737–43.
10. Gleim GW, Stachenfeld NS, Nicholas JA. The influence of flexibility on the economy of walking and jogging. *J Orthop Res*. 1990; 8(6):814–23.
11. Greenspan DS, Pasquinelli AE. *Bst*UI and *Dpn*II RFLPs at the *COL5A1* gene. *Hum Mol Genet*. 1994;3(2):385.
12. Hakim AJ, Cherkas LF, Grahame R, Spector TD, MacGregor AJ. The genetic epidemiology of joint hypermobility: a population study of female twins. *Arthritis Rheum*. 2004;50(8):2640–4.
13. Jones AM. Running economy is negatively related to sit-and-reach test performance in international-standard distance runners. *Int J Sports Med*. 2002;23(1):40–3.
14. Joyner MJ, Coyle EF. Endurance exercise performance: the physiology of champions. *J Physiol*. 2008;586(1):35–44.
15. Lahiri K, Nurnberger JI. A rapid non-enzymaticmethod for the preparation of HMW DNA from blood for RFLP studies. *Nucleic Acids Res*. 1991;19(19):5444.
16. MacArthur DG, North KN. Genes and human elite athletic performance. *Hum Genet*. 2005;116(5):331–9.
17. Maes HH, Beunen GP, Vlietinck RF, et al. Inheritance of physical

fitness in 10-yr-old twins and their parents. *Med Sci Sports Exerc*. 1996;28(12):1479–91.
18. Mokone GG, Schwellnus MP, Noakes TD, Collins M. The *COL5A1* gene and Achilles tendon pathology. *Scand J Med Sci Sports*. 2006;16(1):19–26.
19. Posthumus M, September AV, O'Cuinneagain D, van der Merwe W, Schwellnus MP, Collins M. The *COL5A1* gene is associated with increased risk of anterior cruciate ligament ruptures in female participants. *Am J Sports Med*. 2009;37(11):2234–40.
20. Saunders CJ, Xenophontos SL, Cariolou MA, Anastassiades LC, Noakes TD, Collins M. The bradykinin beta 2 receptor (BDKRB2) and endothelial nitric oxide synthase 3 (NOS3) genes and endurance performance during Ironman Triathlons. *Hum Mol Genet*. 2006;15(6):979–87.
21. Schwarze U, Atkinson M, Hoffman GG, Greenspan DS, Byers PH. Null alleles of the *COL5A1* gene of type V collagen are a cause of the classical forms of Ehlers–Danlos syndrome (types I and II). *Am J Hum Genet*. 2000;66(6):1757–65.
22. September AV, Cook J, Handley CJ, Van der ML, Schwellnus MP, Collins M. Variants within the *COL5A1* gene are associated with Achilles tendinopathy in two populations. *Br J Sports Med*. 2009; 43(5):357–65.
23. Wenstrup RJ, Florer JB, Brunskill EW, Bell SM, Chervoneva I, Birk DE. Type V collagen controls the initiation of collagen fibril assembly. *J Biol Chem*. 2004;279(51):53331–7.
24. Wenstrup RJ, Florer JB, Davidson JM, et al. Murine model of the Ehlers–Danlos syndrome. col5a1 haploinsufficiency disrupts collagen fibril assembly at multiple stages. *J Biol Chem*. 2006; 281(18):12888–95.
25. Wenstrup RJ, Florer JB, Willing MC, et al. *COL5A1* haploinsufficiency is a common molecular mechanism underlying the classical form of EDS. *Am J Hum Genet*. 2000;66(6):1766–76.